SELF-ASSEMBLY OF NUCLEOBASE ANALOGS IN WATER: SUPRAMOLECULAR POLYMERS, CONTROLLABLE MATERIALS AND MODELS FOR PROTO-NUCLEIC ACIDS

A Dissertation
Presented to
The Academic Faculty

by

Brian Joseph Cafferty

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
August 2015

SELF-ASSEMBLY OF NUCLEOBASE ANALOGS IN WATER: SUPRAMOLECULAR POLYMERS, CONTROLLABLE MATERIALS AND MODELS FOR PROTO-NUCLEIC ACIDS

Approved by:

Dr. Nicholas V. Hud, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Gary B. Schuster, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Loren D. Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Adegboyega K. Oyelere
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Date Approved: May 19, 2015

ACKNOWLEDGEMENTS

I express my deepest gratitude to my advisor, Nicholas Hud. Nick provided the ideal mix of intellectual support, creativity, freedom to satisfy my own curiosities, and an environment that has allowed me to grow as an independent researcher. An exceptional teacher and mentor, Nick's insight, guidance and enthusiasm made the work presented here possible, and truly rewarding. Thank you Nick, I hope this is just the beginning of a long and productive relationship.

I thank my advisor Gary Schuster for providing support, a space to work, and insightful discussions about research and what it takes to be a successful researcher. He has taught me to always keep the grand chemical challenges in mind but never overlook the minutiae. Additionally, I greatly acknowledge the rest of my committee: Loren Williams, Yomi Oyelere and Roger Wartell. I've enjoyed our discussions and prospered from your assistance on projects over the years. I am truly grateful for your friendship.

I would like to thank my collaborator, Ram Krishnamurthy, for our many fruitful conversations, his careful analyses of manuscripts, and unique ability to identify and propose experiments to answer important questions. Additionally, I thank Drs. Joshy Joseph and Elizabeth Kuruvilla for oligonucleotide synthesis, Dr. Anton Petrov for modeling expertise, Dr. Keunsoo Kim for monomer synthesis, Dr. Les Gelbaum for NMR assistance, Prof. Lawrence Bottomley for use of AFM, Dr. Jay Forsythe for mass spec analysis, as well as Dr. Miguel Palaez-Fernandez and Prof. Alberto Fernandez-Nieves for help with rheology.

I am in debt to my fellow lab mates, both past and present. Aaron Engelhart was an inspiring and extremely knowledgeable presence during my first two years as a graduate student and I sincerely thank you for your guidance. My colleagues, and friends, Isaac Gállego, Michael Chen and Suneesh Karunakaran were excellent sources of knowledge and creativity; their contributions were essential to the development of numerous projects described in this document. I truly hope that we stay in touch on both a scientific and personal level. I additionally wish to thank David Fialho, Irena Mamajanov, Catarina Musetti, Chao Li, Jaheda Khanam, Denise Enekwa, Rekha Avirah, Adriana Lozoya-Colinas, Ragan Buckley, Tim Lenz, Brandon Laughlin, Gary Newnam, Ford Lannan, Sara Walker, Katie Farley, Linda Nhon, Chelsea Walker, Yesmalie Aleman, Julie Spector-Sprague and Shreyas Athavale. You have all made a major impact on a personal level, as well as on the work presented herein. Thank you.

I greatly acknowledge the CCE for providing me with numerous opportunities during my graduate training. Christine Conwell deserves special thanks for motivating me to take full advantage of these opportunities. Sue Winters, thank you for always being on top of your game as a research administrator and beer connoisseur (and thanks for the homebrew!). I thank NASA, NSF and IBB for funding.

I am thankful for the love and support of my girlfriend, Ashley Allen, for my mother and father's encouragement, my brother Lee for inspiration and my sister Alicia's wisdom.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

A	Adenine
Å	Ångstrom
AFM	Atomic Force Microscope
AMBER	Assisted Model Building with Energy Refinement
AMP	Adenosine Monophosphate
arb. units	Arbitrary Units
AU	Absorbance Units
BA	Barbituric Acid
bp	Base Pair
BMP	5-Ribofuranosyl-barbiturate-5'-monophosphate
C	Cytosine
CAR	1-β-Ribofuranosyl-cyanurate
°C	Degrees Celsius
C18	Octadecyl Carbon Chain Bound Silica
CD	Circular Dichroism
cm	Centimeter
Cy	Cyanuric Acid
CyCo4	Butyric Acid *N*-Substituted Cyanuric Acid
CyCo6	Hexanoic Acid *N*-Substituted Cyanuric Acid
d	Deoxyribo-
D	Dextrorotatory
DAP or D	4,6-Diaminopurine
DMSO	Dimethyl Sulfoxide

DNA	Deoxyribonucleic Acid
D_2O	Deuterium Oxide
E	Energy
ε	Extinction Coefficient
EDTA	Ethylenediaminetetraacetic Acid
ESI	Electrospray Ionization
EtBr	Ethidium Bromide
FNA	Flexible Nucleic Acid
g	Gas
G	Guanine
ΔG	Change in Gibbs Free Energy
G′	Shear Storage Modulus
G″	Shear Loss Modulus
ϑ	isoGuanine
GMP	Guanosine Monophosphate
GNA	Glycol Nucleic Acid
h	Hour
H	Enthalpy
^1H	Proton (in NMR)
HPLC	High-Performance Liquid Chromatography
HRMS	High-Resolution Mass Spectrometer
I	Hypoxanthine
IE	Ion Exchange
ig	Pertaining to a Glycol Nucleic Acid Analog
isoGNA	Glycol Nucleic Acid Analog

K	Kelvin
kcal	Kilocalorie
K_{sp}	Solubility Product Constant
L	Levorotatory
λ	Wavelength
LCMS	Liquid Chromatography-Mass Spectrometry
LNA	Locked Nucleic Acid
M	Molar
$[M]^-$	Molecular Anion
$[M]^+$	Molecular Cation
MA	Melamine
MAC	Minimal Assembly Concentration
MD	Molecular Dynamics
mdeg	Millidegrees
MHz	Megahertz
min	Minute
µL	Microliter
mm	Millimeter
MM	Mismatch
µm	Micrometer
µM	Micromolar
MMP	Melamine-Riboside Monophosphate
m/z	Mass-to-Charge Ratio
NFA	Normalized Fraction Assembled
NH_4OAc	Ammonium Acetate

NMR	Nuclear Magnetic Resonance
nm	Nanometer
nM	Nanomolar
NOE	Nuclear Overhauser Effect
nt	Nucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
pH	Measurement of Hydrogen Ion Concentration
pK_a	Acid Dissociation Constant
POM	Polarized Optical Microscopy
ppm	Parts Per Million
PRF	Proflavine
ps	Picosecond
r	Ribo-
rel. Abs.	Relative Absorbance
RNA	Ribonucleic Acid
ROE	Rotating Frame Nuclear Overhauser Effect
RP	Reverse Phase
RT	Reverse Transcriptase
S	Entropy
sec	Second
T	Thymine
TAP	2,4,6-Triaminopyrimidine
TARC	5-β-Ribofuranosyl-2,4,6-Triaminopyrimdine
TEA	Triethylamine

TEAA	Triethylammonium Acetate
TEM	Transmission Electron Microscopy
T_m	Melting Temperature
TNA	Threose Nucleic Acid
TSP	Trimethylsilyl-2,2,3,3-Tetradeuteropropionic Acid
U	Uracil
UNA	Unlocked Nucleic Acid
UMP	Uridine Monophosphate
UV-Vis	Ultraviolet-Visible
X	Xanthine

SUMMARY

Uncovering the origin of RNA is essential for understanding the origins of life. The persistent inability of chemists to identify a plausible prebiotic route to RNA polymers, along with the seemingly optimal structure of RNA for its functions in extant life, argue in favor of the hypothesis that RNA is a product of chemical or biological evolution. To understand the origin of RNA, we must consider which molecules could have originally acted in place of RNA's substructures in the oldest ancestor of RNA (or proto-RNA). Major challenges to uncovering the chemical structure of proto-RNA include finding molecules that would have spontaneously undergone molecular selection and covalent assembly into nucleotides and RNA-like polymers, within the complex mixture of the "prebiotic soup" and without the aid of enzymes.

Towards overcoming these challenges, I have investigated alternative heterocycles as possible nucleobases of an ancestral form of RNA. Several plausible proto-nucleobases have been identified that are capable of self-assembly in water and facile nucleosidation reactions. Nucleosides and nucleotides are spontaneously generated in high yield when 2,4,6-triaminopyrimidine, melamine or barbituric acid are mixed with ribose or phosphorylated ribose in water. The complementary heterocycles, nucleosides and nucleotides self-assemble through Watson-Crick-like base pairing interactions, demonstrating a prebiotic mechanism of organization that might overcome problems associated with base selection and the spatial ordering of nucleosides and nucleotides before the advent of polymerases.

Throughout this dissertation I will also discuss how these investigations have provided general insight into molecular self-assembly in water, an essential process underlying developmental strategies for responsive and potentially therapeutic materials. Hexameric rosette-based hydrogelation motifs have been identified that possess highly configurable architectures, in part due to the chemical simplicity of their monomers. Because the heterocycles used in these systems resemble the RNA nucleobases in structure and interactions, the free energy of monomer associations could be estimated, providing a greater understanding of the interactions that can be employed by monomers in aqueous solution. The hydrophobic effect was found to drive assembly of the water-soluble monomers into long noncovalent structures through the intermediate formation of large hydrophobic rosette surfaces, resulting in the observed cooperative assembly mechanism. For one system, monomer assembly is shown to exhibit the theoretical limit for sensitivity to pH change for a supramolecular polymerization system. A new strategy is also demonstrated to further control the supramolecular polymerization process by employing planar molecules, heterocycles we have termed *noncovalent* terminators, that stack on the rosette surface and can be used to precisely tune the length of supramolecular polymers and the bulk properties of hydrogels. These findings demonstrate that significant contributions to both basic and applied chemistry can directly result from searching for key questions in the origin of life.

In the final chapters I will discuss studies involving modified nucleic acid backbones, which have led to the development of a DNA-based, and polymerase compatible, dynamic combinatorial chemistry system and provided insight into the contribution of the backbone to the binding of intercalators to nucleic acid duplexes.

CHAPTER 1

INTRODUCTION[a]

The chemist finds illustration, inspiration, and stimulation in natural processes, as well as confidence and reassurance since they are proof that such highly complex systems can indeed be achieved on the basis of molecular components.

Jean-Marie Lehn, From Matter to Life: Chemistry?! (1995)

1.1 Background

Nucleic acids were first discovered by Friedrich Miescher in 1869; however their simple chemical composition led many to rule out any potential genetic role. It was not until the next century, following seminal studies by Avery et. al. [1], and later Hershey and Chase [2], that the nucleic acids were experimentally verified to support an informational function. The molecular basis for how nucleic acids store and transfer genetic information was revealed when the double helix structure was solved by Crick and Watson in 1953 [3]. Their model provided an elegant demonstration of the interplay between molecular structure and function, and helped usher in a golden age in molecular biology. Discoveries made over the next two decades prompted Crick to propose the *Central Dogma of Molecular Biology*, which describes the flow of information from DNA to RNA to proteins [4]. As the molecular foundation of biology began to reveal itself, so too did clues regarding its distant past and a new optimism to uncover its origin.

a. This chapter was adapted from previously published work and is reproduced with permission. Cafferty, BJ; Hud, NV. "Was a pyrimidine-pyrimidine base pair the ancestor of Watson-Crick base pairs? Insights from a systematic approach to the origin of RNA" Isr. J. Chem. (2015). Cafferty, BJ; Hud, NV. "Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology" Chem. Biol. (2014). Hud, NV; Cafferty, B.J.; Krishnamurthy, R; Williams, LD. "The origin of RNA and "My grandfather's axe" Chem. Biol. (2013).

Living organisms are dependent on a synergistic cooperation of polymers and multimolecular assemblies (i.e, nucleic acids, proteins, polysaccharides, lipid membranes). It is difficult to imagine contemporary life without the participation of any one class of extant biopolymer. However, such a complex system containing a diverse set of polymer types raises a causality dilemma at the origin of life, since proteins perform the overwhelming majority of enzymatic function (including nucleic acid synthesis) and the nucleic acids are the polymer class that stores the information that encodes proteins, it appears improbable to have one polymer type form spontaneously without the aid of the other. This, in part, has led many researchers to embrace the RNA World Hypothesis [5], which proposes a period when RNA performed both information storage and catalysis. This hypothesis received additional support with the discovery of catalytic RNAs in the 1980's [6-9], and the subsequent finding that ribosomal peptidyl transfer result from an RNA catalyst [10, 11]. While the RNA World Hypothesis simplify life's chemical origins by reducing the number of participating polymers, there is little reason to believe that polysaccharides, lipid membranes, and uncoded polypeptides did not simultaneously emerge along with nucleic acids. This notwithstanding, the origin of RNA is of central importance to understanding the origin of life.

The nucleotides of RNA appear to be an amalgam of three substructures, or building blocks, that have their origins in three distinct starting materials (Figure 1.1). The nucleobases (e.g., adenine, guanine, cytosine and uracil) are heterocycles that look like molecules derived from hydrogen cyanide (HCN) or closely related organic compounds (e.g., formamide, ammonium formate, urea). Oró discovered in 1960 that adenine forms spontaneously in concentrated solutions of ammonium cyanide [12].

Ribose, a carbohydrate (chemical formula $C_5H_{10}O_5$), appears to be the product of formaldehyde (H_2CO) polymerization. Bulterow reported in 1861 that sugars are formed by the polymerization of formaldehyde in the formose reaction, a reaction that produces ribose as a minor component among a plethora of other sugars [13]. Phosphate is an inorganic molecule abundant on the Earth's surface, however it likely was not in a soluble state as it readily forms insoluble minerals with divalent cations (i.e., Ca^{2+}, Fe^{2+}, Mg^{2+}).

The three substructures of RNA are joined to each other by chemical bonds that are, in principle, the result of *condensation-dehydration* reactions; reactions that produce a water molecule upon bond formation (Figure 1.1). Given this modular structure of RNA, chemists have long hypothesized that the substructures of RNA were first produced by three distinct reactions, and then connected by a sequence of reactions that generated nucleosides, then nucleotides, and finally RNA polymers. However, the formation of RNA under prebiotic conditions has not been successfully demonstrated even after 50 years of directed research.

In the early 1970s Orgel initiated the quest to produce RNA from its pre-formed substructures by drying and heating the free nucleobases with ribose. Early results were encouraging. In particular, β-adenosine (the ribose nucleoside of adenine) was produced in modest yields (up to 3%) [14]. However, prebiotic reactions have never been established that significantly increase the yield of adenosine. Furthermore, the other canonical nucleobases of RNA and DNA (guanine, uracil, cytosine and thymine) do not form their respective nucleosides in detectable yields in analogous reactions [15]. Frustrated by the inability to produce nucleosides by simple drying reactions, some

chemists began to refer to this putative step in prebiotic RNA formation as *'the nucleoside problem*.' The additional challenges of phosphorylating nucleosides to make nucleotides, and polymerizing nucleotides to create polymers, has been collectively referred to as *'the water problem'* [16], to emphasize that the formation of these bonds in water is thermodynamically unfavorable; a vexing problem for prebiotic chemists.

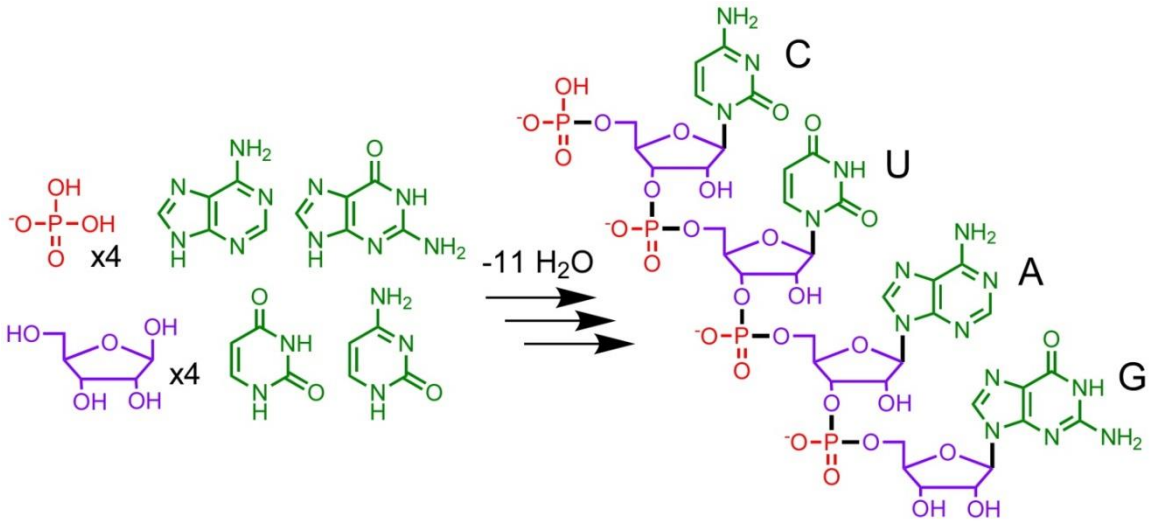


Figure 1.1 Chemical structure of RNA and RNA building blocks. RNA is composed of phosphate, ribose and four nucleobases (adenine, guanine, uracil and cytosine). Each building block is incorporated into RNA through a condensation-dehydration reaction. The above RNA polymer requires 11 condensation-dehydration reactions between the building blocks shown, arrows indicate multiple reactions. Bonds formed are shown with bold black lines. Note that only some of these reactions have been demonstrated in model prebiotic experiments. Reprinted with permission from reference 23.

Prior to the development of modern analytical instrumentation, prebiotic chemists primarily focused on demonstrating that it is possible to abiotically produce a particular biological building block (for which authentic standards were available). Such studies demonstrated that a variety of starting materials and reaction conditions produce the canonical bases of RNA. The development of techniques that allow the identification of molecules within complex mixtures (e.g. LCMS) has now revealed that model prebiotic reactions that start with even just one or two organic compounds can produce a substantial number of heterocycles, including molecules that are closely related to the canonical nucleobases [17, 18]. Thus, while the RNA nucleobases are now widely accepted as being prebiotic molecules, the production of complex mixtures by even the simplest model prebiotic reactions necessarily shifts attention to other questions regarding the prebiotic nature of the nucleobases. Including, *How?*, *When?*, and *Why?* were the canonical nucleobases first selected from a complex pool of heterocycles (with similar chemical and physical properties) for incorporation into nucleic acids? Possible answers to these questions are discussed below.

The prebiotic synthesis of ribose is more troublesome than the synthesis of the nucleobases. The small yields of ribose by the formose reaction (often cited as a possible prebiotic source of sugars) and the relative instability of ribose, even in comparison to other sugars, have been cited as reasons to doubt that ribose was part of the first nucleic acids [19]. Benner and coworkers have attempted to circumvent these problems using borate minerals. These researchers have demonstrated that borate slows ribose degradation by chelation of *cis*-diols, and increases ribose production in the formose reaction by preferentially favoring pathways that produce a subset of possible sugars,

including ribose [20]. Their chemical mapping of the complex formose reaction under the influence of borate is impressive. However, geologists have questioned the prebiotic relevance of borate in prebiotic reactions on the grounds that borate minerals would not have existed on the early earth [21]. Nevertheless, the observation that borate can selectively enhance the products of a model prebiotic reaction emphasizes the importance of considering how inorganic (and organic) co-solutes could have influenced the contents of the prebiotic chemical inventory.

These challenges, juxtaposed that the seemingly optimal structure of RNA (i.e., with regards to its functionality in extant life), have led to the hypothesis that RNA is a product of chemical or biological evolution, originating from a chemically distinct informational polymer [22]. This hypothetical first genetic polymer is referred to as proto-RNA, with the understanding that the evolution of proto-RNA to extant RNA could have transpired through an unknown number of intermediate pre-RNAs (Figure 1.2) [22]. The structure of proto-RNA and pre-RNAs is as much a historical question as it is a chemical question. While time has erased virtually all physical traces of the molecules and processes that gave rise to nucleic acids, many prebiotic chemists remain optimistic that fundamental chemical studies —informed by chemical clues in extant life and constrained by knowledge of the early Earth— will eventually allow us to discover molecules and reactions with the ability to spontaneously provide the polymers from which life can emerge [23].

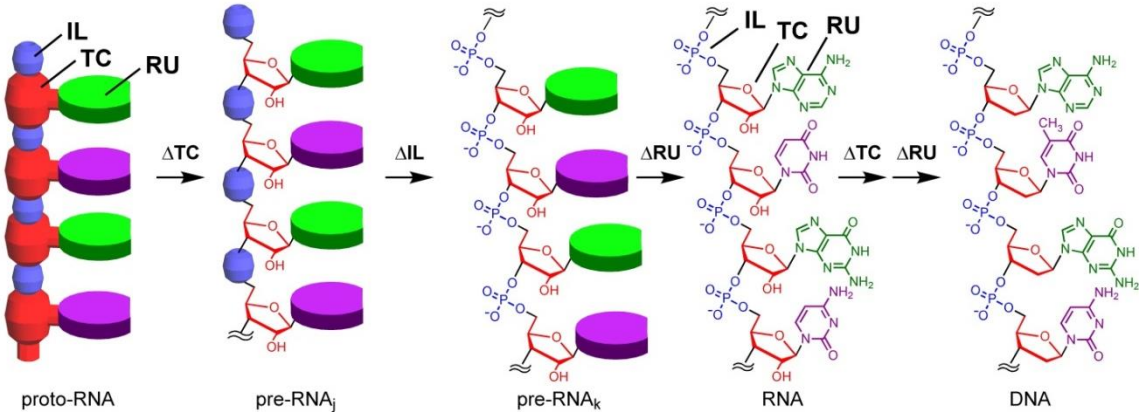


Figure 1.2 Schematic representing of a hypothetical evolutionary lineage of nucleic acids from proto-RNA to RNA and DNA. Intermediates between proto-RNA and RNA are shown for illustrative purposes only, and are not intended to imply that changes at any particular position proceeded in the particular order or number of steps shown. RUs indicates recognition unit (e.g., nucleobases), TC is trifunctional connector (e.g., ribose) and IL is ionic linker (e.g., phosphate). Reprinted with permission from reference 22.

Demonstrating how the first building blocks were selected and covalently linked together to form a polymer capable of hereditary is necessary for the realization of most models concerning the origin of informational polymers. This research objective requires the analysis of molecules that were likely present on the prebiotic earth and have the potential to self-assemble into the genetic polymer that ultimately allowed the emergence of life. As the nucleobases code genetic information and direct molecular selection, investigating molecules that may have functioned as proto-nucleobases provides a key starting point to direct research aimed at uncovering the structure of proto-RNA nucleotides and polymers. This chapter begins with a discussion of what was likely required by a heterocycle to function as a proto-nucleobase, followed by a review of the efforts to identify the ancestral nucleobases of the earliest pre-RNAs, as well as related studies that have provided valuable insights, and ends with a discussion of the general

properties that enable, or govern, self-assembly of small molecules in water and models for the spontaneous formation of informational polymers.

1.2 Considerations for the Selection of Nucleobases for a Proto-RNA Polymer

1.2.1 Why *Not* the Extant Nucleobases from the Beginning?

The base pairs of RNA and DNA are now iconic symbols of molecular recognition. Few rules in biochemistry are better defined than the pairing of adenosine with thymine and guanine with cytosine. In addition to the complementary hydrogen bonding of the bases, the stacking of paired bases may represent one of the densest possible forms of information storage. Moreover, the pseudo dyad of the Watson-Crick base pairs allows any nucleotide sequence to exist as a DNA double helix with the same helical structure. Taken together, it would seem that A, T, G and C are obvious selections for DNA. In light of the popular RNA world hypothesis, it is often assumed that the bases of DNA are largely a carryover from RNA, with the substitution of T for U being the only difference between the canonical nucleobases of RNA and DNA. In its role as the long-term repository of genetic information, it is understandable why DNA would have substituted T for U (e.g., to retain the identity of a G•C base pair after deamination of C). Just as many origins researchers are willing to accept that U came before T, and that U was replaced by T for functional reasons, we should also be open to the possibility that the extant bases of RNA have evolved from other heterocycles. There are multiple reasons why the extant nucleobases appear to be optimal for extant life, but, as discussed

below, these same nucleobases are extremely problematic when it comes to the prebiotic origins of RNA.

Glycosidic bond formation — The hydrolytic stability of the bond between the nucleobases and ribose (i.e., the glycosidic bond) make the extant nucleobases good choices for use in the storage and transfer of genetic information within biological organisms [24, 25]. Conversely, the formation of the glycosidic bond between a free nucleobase and ribose has proven to be one of the most vexing challenges to the prebiotic synthesis of RNA. The relative ease with which heterocycles can be formed from a number of starting reagents (e.g., HCN, formamide, urea) [12, 17, 18], as well as possible prebiotic routes to a multitude of sugars [13, 26, 27], supports the long-standing proposal that the nucleosides of the first nucleic acids could have been formed by the condensation of pre-existing bases and pre-existing sugars. As briefly stated above, experiments by Orgel and co-workers demonstrated that adenine will form adenosine in modest yields when dried and heated with ribose [14, 28]. In contrast, analogous prebiotic reactions for the joining of the extant pyrimidine bases with ribose have never been found, and early reports that guanosine can be made by drying guanine with ribose appear questionable [28]. If the first informational polymers of life self-assembled without the aid of enzymes or chemical activation, there might have been proto-nucleobases that more easily formed nucleosides with ribose, or another trifunctional linker (i.e., a proto-ribose). If this was the case, then the first glycosidic linkages (or analogous linkages) would likely have been less kinetically stable, following what is expected for a bond with a lower transition energy barrier, or what can be thought of as an ‘easy on, easy off’ rule for nucleoside-formation reactions. A good example is inosine, a nucleoside that can be formed by

drying its free nucleobase with ribose [14, 28], but whose glycosidic bond is much more susceptible to hydrolysis when compared to the canonical nucleosides [24].

The inability to form nucleosides with uracil and cytosine in simple drying-heating reactions motivated Orgel to propose that the cytidine base might have originally been formed piecewise on a sugar scaffold (thereby circumventing the need for glycosidic bond formation) (Figure 1.3) [29]. The feasibility of this approach was more recently demonstrated by Powner et al. who showed that cytidine could be synthesized from small, plausibly prebiotic molecules, albeit in a multistep reaction that required the ordered addition of reagents and purification of an intermediate [30]. Given the prebiotic challenges associated with such reactions [16], and the possibility that other bases came before the pyrimidines, the propensity for a heterocycle to form a glycosidic bond with ribose (or other sugars) in a model prebiotic reaction can be considered an indicator of its viability as a proto-nucleobase.

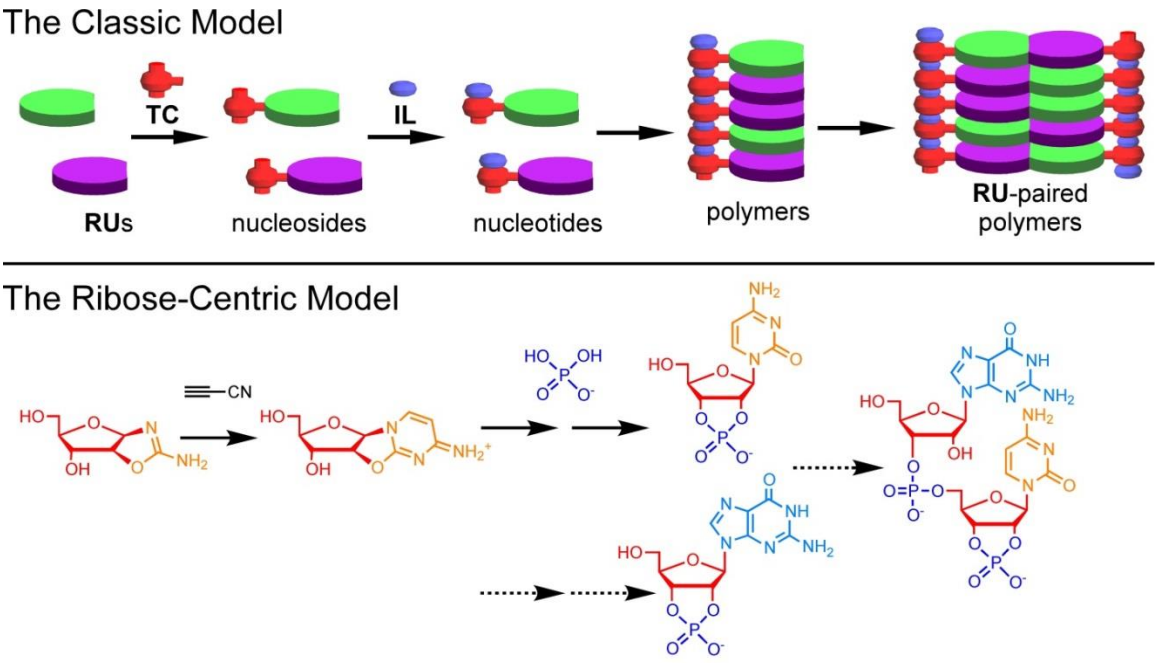


Figure 1.3 Proposed models for the origin of RNA and pre-RNA. The Classic model involves the trifunctional connector (TC), ionic linker (IL) and recognition unit (RU) assembling sequentially to produce nucleotides (or proto-nucleotides) before becoming opolymerize dto form RNA (or proto-RNA). The ribose centric model which was originally proposed by Orgel, shows the cytosine base being built on a preformed sugar scaffold. Like the Classic Model nucleotides are formed prior to polymerization. Reprinted with permission from reference 22.

The paradox of base pairing — It has been known for decades that the free nucleobases of RNA (and their nucleosides) do not form Watson-Crick base pairs in water [31]. The hydrogen-bonding tendency of these heterocycles is easily satisfied by water, which is approximately 100 M in hydrogen bond donors and acceptors (i.e., oxygen lone pairs). Thus, in water the nucleobases tend to form columnar stacks, which sequester their hydrophobic surfaces, while presenting their Watson-Crick edges to solution. Hud and co-workers have proposed that the non-pairing of the extant

nucleobases and their mononucleotides in aqueous solution presents a "Paradox of Base Pairing" with respect to the origin of RNA [32, 33]. Model prebiotic reactions have revealed that the abiotic reactions that first produced the nucleobases would have likely produced many more heterocycles than just those found in extant RNA as well as a plethora of other molecules that would have similar (or greater) reactivity (Figure 1.4). The complex mixture of the 'prebiotic soup' and the non-pairing of the free nucleobases of RNA thereby presents a paradox: *If the free bases of RNA were not pairing before they were attached to a polymer backbone, then how would life (or prebiotic chemistry) have 'known' to select and incorporate these bases into polymers with the anticipation that they would form base pairs in the polymeric state?* Evolution does not predict what will be advantageous in the future. Thus, in the absence of a mechanism for the selection of pairing bases, the spontaneous polymerization of heterocycles would have resulted in the production of polymers with many non-pairing bases; a situation that could have wasted valuable resources and thwarted the emergence of pairing nucleic acid polymers. As a possible resolution to this paradox, the first bases of proto-RNA may have either been different from those found in life today (with the ability to pair as monomers in aqueous solution), or other molecules may have been present that facilitated the selection of pairing bases [32, 34, 35]. Below, additional heterocycles are considered that could have circumvented this paradox.

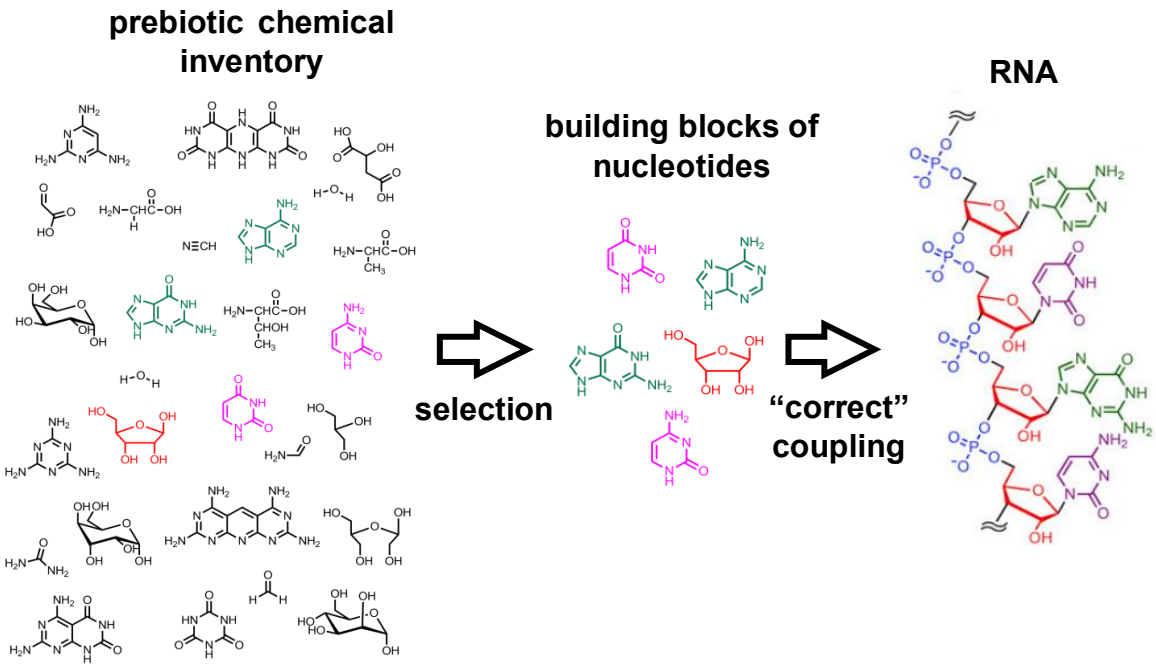


Figure 1.4 Schematic showing obstacles facing the spontaneous formation of RNA. These include the selection of building blocks from the hypothetical complex prebiotic environment and specifically linking these building blocks together in way to enable/maintain function.

1.2.2 Identifying Candidates for the Proto-RNA Nucleobases

The question of why nature chose the extant nucleobases will likely prove to be easier to answer than why nature chose ribose. The impact of altering the structure of ribose within the backbone of RNA is, arguably, less predictable than the impact of making changes to the nucleobases. As discussed below, some changes in sugar structure result in decreased duplex stability, while others lead to increased duplex stability [36]. In contrast, it is easy to imagine that changing any of the functional groups on the extant nucleobases would have a predictable impact on the stability of a Watson-Crick base pair. Furthermore, sugars are highly soluble in water and possess similar functional groups (i.e., primary and secondary alcohols). It is therefore not obvious how sugars would have been selected at the monomer level, prior to the emergence of enzymes. In contrast, the physical and chemical properties of even closely related heterocycles can vary tremendously (see below).

Any proposal for the identity of the building blocks of proto-RNA must take into consideration both availability and functionality. That is, candidate molecules must be prebiotically plausible (i.e., formed through abiotic reactions that do not require biological intervention), environmentally available (e.g., soluble, sufficiently reactive yet stable) and chemically functional (e.g., able to function in a similar capacity as their analogous components of extant RNA). As mentioned above, to overcome the chemical and selection problems that face abiotic incorporation of the contemporary nucleobases into RNA, proto-nucleobases may have readily formed glycosidic bonds and facilitated their own mutual selection from complex mixtures. The latter requirement may include the ability of the earliest nucleobases to direct the formation of supramolecular structures

that organized these molecules (or their proto-nucleotides) in a manner that promoted their covalent linkage into proto-RNA polymers.

Changes to both the backbone and the nucleobases would have occurred in a manner that allowed the transfer of information between successive generations of pre-RNA polymers, a supposition that is part of the more general 'principle of continuity' that Orgel argued must be satisfied when considering potential prebiotic ancestors of RNA [37]. Given that genetic information in extant life is transferred through Watson-Crick base pairing (in DNA replication, RNA transcription, and RNA translation) it is parsimonious (or at least convenient) to hypothesize that even the earliest nucleobases of pre-RNA possessed hydrogen-bonding edges that would have been compatible with the hydrogen bonding patterns of Watson-Crick base pairs.

In addition to possessing particular hydrogen bonding edges, the most likely candidates for the proto-RNA bases would be those that also place their glycosidic bonds at relative distances and angles that are the same (or very close) to those found in RNA today. Maintaining this geometry would have, to some extent, preserved secondary and tertiary structures as proto-RNA evolved into successive pre-RNAs, and eventually into extant RNA.

1.2.3 The Challenge and Benefits of Alternative Nucleobases

Leslie Orgel called the possibility that RNA evolved from another polymer a "gloomy prospect" [38], because if the first nucleic acids were different from those of extant life we are then faced with sorting through a seemingly enormous number of molecules to find those that could have given rise to proto-RNA. Over the past twenty-

five years several research groups, beginning with Benner and colleagues [39-41], demonstrated that numerous alternative heterocycles are able to substitute for the extant nucleobases in RNA or DNA duplexes. In these experiments the non-natural base pairs were typically flanked by standard Watson-Crick base pairs, thereby testing the compatibility of alternative nucleobases with extant nucleic acid helical structure, which may be more restrictive than backbone structure. Just as ribose can be replaced by alternative trifunctional linkers in the nucleic acid backbone without loss of base pairing [36], these base-substitution studies have shown that the canonical bases are not unique in their ability to form stable duplexes with base pairs.

Given that the components of extant RNA are not unique in their ability to form duplex structures or to even support evolution [42-44], we are truly faced with Orgel's gloomy prospect of sifting through a large number of candidate molecules to find the constituent building blocks of proto-RNA. However, all hope is not lost! A systems-level approach for the identification of candidate proto-RNA nucleobases may be carried out by looking for: 1) plausible proto-RNA nucleobases that are in the same *chemical space* that contains the extant nucleobases, and 2) nucleobases that can circumvent the multiple problems that have stymied the spontaneous formation of extant RNA in model prebiotic reactions.

1.2.4 The Chemical Space: Purines, Pyrimidines and Triazines

Eschenmoser once stated that his search for alternatives to ribose was limited to other sugars that could be synthesized by the same chemistry (e.g., from formaldehyde) [36]. Likewise, a search for candidate proto-nucleobases can be focused initially on

heterocycles that utilize the same, or similar, scaffolds as the extant bases. As a first pass, all 81 purines and pyrimidines that have H, NH_2 or O as exocyclic groups are shown in Figure 1.5. This restriction can be partially justified by parsimonious arguments of chemistry. Simple and plausibly prebiotic organic molecules, such as formamide and urea [17, 18], have been shown to form a wide assortment of purines and pyrimidines. Thus, numerous purines and pyrimidines distinct from the extant nucleobases would have likely been present at the time of the earliest abiotic synthesis of nucleic acids.

Prebiotic heterocycles could have also been present that contained different heteroatoms. For example, 2-thiouridine, a nucleobase found in the tRNAs of all three branches of life, has been proposed as a potential prebiotic nucleobase that would have enhanced the fidelity of template-directed synthesis by suppressing formation of G•U wobble base pairs [45]. Additionally, the s-triazines are attractive alternatives as they are structurally related to the pyrimidines, and some have been found alongside pyrimidines and purines in meteorites [46, 47] and in model prebiotic reactions [18]. Triazines are also capable of making Watson-Crick-like base pairing and have been proposed to constitute pre-RNA [48, 49]. For these reasons triazines should also be considered.

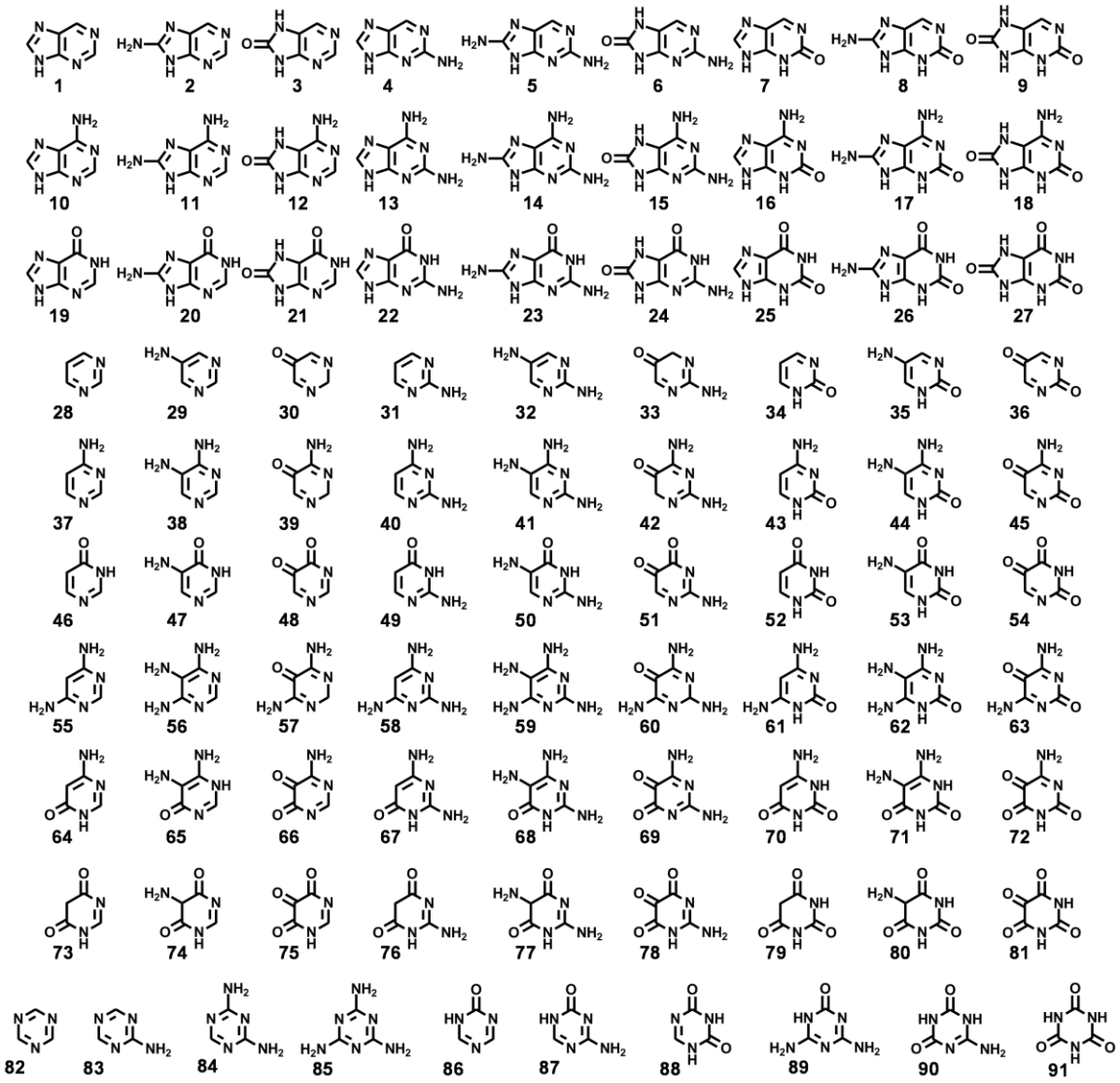


Figure 1.5 The chemical space defined by heterocycles with either the purine, pyrimidine or triazine ring systems and the exocyclic groups O, NH_2 or H. The purine class contains 27 different heterocycles with substitutions at the 2, 6 and 8 positions. The pyrimidine class contains 54 different heterocycles with substitutions at the 2, 4, 5 and 6 positions. The triazine class contains 10 different heterocycles with substitutions at the 2, 4, and 6 positions. For clarity, 12 pyrimidines are not shown that contain the same exocyclic substitutions as **36**, **45**, **48**, **51**, **54**, **63**, **66**, **69**, **72**, **75**, **78**, and **81**, but are in a more reduced form. Reprinted with permission from reference 112.

1.3 Consideration of Alternative Base Pairs in Proto-RNA

1.3.1 Alternative Purine-Pyrimidine Base Pairs

As potentially the smallest possible step back from the extant bases, consider the scenario in which the recognition units of proto-RNA also formed base pairs comprised of one purine and one pyrimidine. This possibility, along with the above-stated requirement that information transfer be forward compatible as proto-RNA evolved to RNA, places substantial constraints on which recognition units could have come before the extant nucleobases. Specifically, we are limited to the purine and pyrimidine bases with hydrogen bond donor and acceptor groups that match the Watson-Crick faces of the extant bases. These edges need not be identical, but each putative proto-nucleobase must be able to form a Watson-Crick base pair with an extant base. To illustrate this point, the extant Watson-Crick base pairs are shown in Figure 1.6 along with three alternative base pairs that satisfy this structural criterion for candidate proto-RNA base pairs. In one case, the H on C2 of adenine is changed to NH_2, becoming 2,6-diaminopurine (D, **13**) (Figure 1.6C). This change is acceptable because the NH_2 group does not interfere with forming a Watson-Crick base pair with uracil (or a proto uracil). D is plausibly prebiotic, being found in some meteorites [50] and synthesized in model prebiotic reactions [51, 52], and has even been shown to enhance non-enzymatic template directed synthesis when used in place of adenine [53]. Moreover, as discussed below, this base looks attractive as an intermediate between the bases of proto-RNA and extant RNA. Nevertheless, it is not clear at this point that D would help resolve other problems associated with proto-RNA formation (i.e., nucleoside formation, the paradox of base pairing).

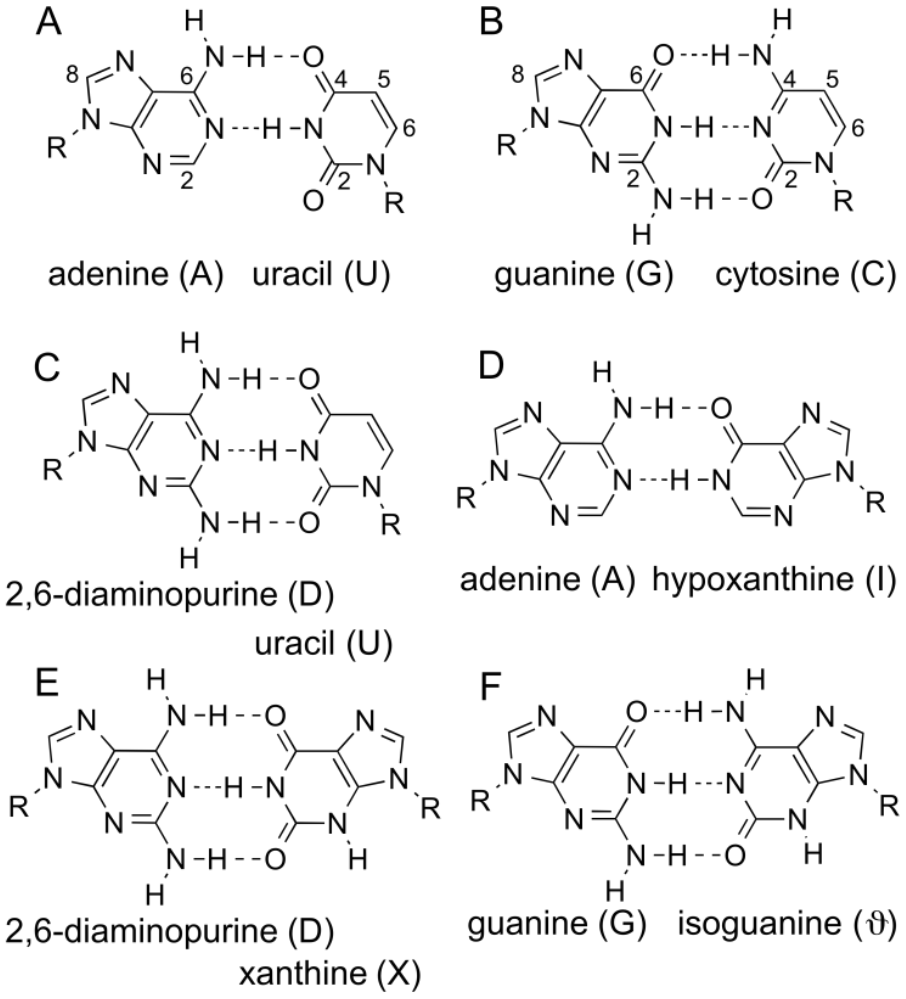


Figure 1.6 Selection of base pairs that have been previously studied. (A-B) Watson-Crick base pairs of the canonical nucleobases. (C-F) Watson-Crick-like base pairs with several non-canonical nucleobases that have been experimentally investigated. Reprinted with permission from reference 112.

The replacement of the NH_2 group on guanine with a H to give hypoxanthine (I, **19**) is also an acceptable path backwards to a potential proto-RNA nucleobase (Figure 1.6D). Hypoxanthine is a purine nucleobase with water solubility that is more than 100-fold greater than that of guanine [54]. Hypoxanthine is also attractive as a possible ancestor of guanine because, as mentioned above, hypoxanthine forms its nucleoside (inosine) in relatively good yield when dried and heated with ribose [14, 28].

Similarly, replacement of the NH_2 with H on cytosine produces 2-pyrimidinone (**34**), a pyrimidine that has also been shown to form nucleosides by drying and heating with ribose in a model prebiotic reaction [55]. While 2-pyrimidinone nucleosides can form a two-hyrogen bonded base pair with guanine, the stability of this base pair is significantly lower than that of a G•C base pair [56]. We note that there is also experimental evidence that 4-pyrimidinone (**46**) also forms nucleosides upon drying and heating with ribose [57], but like its isomer 2-pyrimidinone, this nucleobase is not expected to form a particularly stable Watson-Crick base pair.

While a greater number of non-canonical nucleobases will form base pairs that are compatible with the extant a DNA duplex [41], as mentioned above, we are presently limiting our consideration of candidate proto-nucleobases to those that can form Watson-Crick base pairs with an extant base. As an example of heterocycles ruled out by this criterion, we do not consider isoguanine (**16**) and isocytosine (**49**) (Figure 1.6F), which can form a base pair that is isomorphous with the extant Watson-Crick base pairs [39], but neither heterocycle can form a Watson-Crick base pair with an extant nucleobase.

In addition to the exocyclic groups on the Watson-Crick edges, the exocyclic group on the C8 position of the extant purines and the C5 and C6 positions of the extant

pyrimidines could be changed from H to NH_2 or O without disrupting Watson-Crick base pairing. The prebiotic advantages of making either change to adenine or guanine is not clear. Moreover, replacing H with O at the C8 position of guanine to create 8-oxoguanine (**24**) further decreases solubility in water to levels so low that it is difficult to conduct experiments with this molecule [58]. Changing the 8 position of guanine to NH_2 increases the propensity for self-pairing, which could compete with Watson-Crick pairing, as 8-amino-guanine (**23**) has been shown to form a symmetric hemi-protonated Hoogsteen-Hoogsteen pair [59]. These hypothetical changes highlight problems often associated with candidate proto-nucleobases, i.e., poor solubility would limit nucleobase availability for incorporation into polymers, and a propensity to form self-structures could lead to polymers that lack the ability for high fidelity information transfer.

Finally, efforts using template-directed synthesis reactions as models of RNA replication before the emergence of enzymes have revealed a fundamental problem with the hypothesis that life has always used the canonical purine-pyrimidine base pairs [60-62]. Purine-rich strands do not work well as templates for the synthesis of pyrimidine-rich strands. Two barriers are apparently the main cause of this problem. First, the pairing energy of free pyrimidine nucleotides with purine nucleotides of a template strand is extremely weak. In the case of A•U base pairs, just two consecutive adenine bases on a template strand can be enough to cause truncation of the product strand. Second, while the stronger association of G•C base pairs allows G-rich templates to be read with better efficiency, G-rich templates tend to self-associate by forming G-quadruplex structures. Overall, these template-directed polymerization studies have illustrated the difficulty of

polymerizing cytidine and uridine mononucleotides on purine-rich templates without the aid of enzymes.

1.3.2 Alternative Purine-Purine Base Pairs

Given the challenges associated with starting nucleic acids with purine-pyrimidine base pairs, a number of proposal have been put forth about the possibility that nucleic acids started with either two purines or two pyrimidines. In the 1960s Leslie Orgel and Francis Crick both speculated that nucleic acids once used base pairs different from those of extant RNA. The earliest alternative base pair discussed by Crick, an idea that he credited to Orgel, is the purine-purine base pair A•I (Figure 1.6D) [63]. Prebiotic chemists have long assumed that adenine was one of the earliest molecules of life. It was the first extant nucleobase synthesized in a model prebiotic reaction [12], and it was subsequently shown to be produced by a wide range of plausible prebiotic reactions [18, 52, 64-68]. The observation that hypoxanthine (the base of inosine) is found along with adenine in some meteorites [50], and that these two bases can be produced in the same model prebiotic reactions [52, 69], makes the idea of a proto-RNA with only purine nucleobases appealing.

During the earliest stages of the RNA chemical etiology effort, the Eschenmoser laboratory, and independently Herdewijn and coworkers, explored the base pairing capability of "homo-DNA", a DNA analog containing the larger hexopyranose sugar cycle, i.e., 2',3'-dideoxyglucopyranosyl-(6' $\rightarrow$ 4')-oligonucleotides (or "ddGlc") [70, 71]. Investigations of homo-DNA showed for the first time that Watson-Crick-like pairing between duplexes composed solely of purine-purine base pairs could be achieved [72]. In

addition to A•I base pairs, the oligonucleotides of this study included a variety of sequences with guanine (G, **22**), isoguanine (ϑ, **16**), xanthine (X, **25**), and 2,6-diaminopurine (D, **13**). It was found that D•X and G•ϑ base pairs (Figure 1.6E, 1.6F; respectively) are surprisingly stable when attached to the ddGlc backbone [72]. For example, the duplex formed by two hexamers with six G•ϑ base pairs (i.e., ddGlc(G_6) and ddGlc($ϑ_6$)) has a T_m of 61°C, and a duplex formed by two hexamers with six X•D base pairs (i.e., ddGlc(X_6) and ddGlc(D_6)) has a T_m of 66°C. For comparison, an analogous DNA duplex (e.g., d(G_6) with d(C_6)), under the same conditions, would be expected to have a T_m of less than 30°C [73]. Given that that G•ϑ and D•X base pairs have the same number of hydrogen bonds as a G•C base pair, the enhanced stability of homo-purine ddGlc duplexes is likely due to a greater stacking area of purine-purine base pairs (particularly interstrand stacking) when attached to the ddGlc backbone.

Studies of hexopyranosyl nucleic acids revealed additional factors that must be considered when investigating pre-RNA candidates. For example, a significant challenge to the prebiotic relevance of homo-purine ddGlc oligonucleotides is the rather promiscuous pairing of purine nucleobases. Specifically, there is a tendency of the purine bases to form Hoogsteen-type pairs with themselves, including A•A, D•D, ϑ•ϑ, and G•G base pairs. In fact, the ddGlc A•A homobasic assembly was found to be stronger then the canonical A•T base pair. Again, such self-pairing would complicate information transfer, particularly in the earliest stages of life when proofreading enzymes would have not been available. This problem became even more obvious when the Eschenmoser laboratory investigated true hexopyranosyl sugars (as opposed to deoxy variants) [74]. When both allose and altrose sugars were substitutes for ddGlc, the resulting duplexes had an even

greater preference to use Hoogsteen-type pairs. These findings brought to light the close relationship between backbone structure and preferred nucleobase pairing structure, as some backbones will favor non-informational pairing (e.g., A•A and G•G base pairs) for purine-rich oligos.

After decades of speculation of possible prebiotic A•I base pairs, Battersby et al. were the first to report on the stability of DNA duplexes containing only Watson-Crick-like purine-purine base pairs, including A•I base pairs. To circumvent the problems of low A•I base pair stability, as well as the tendency for G-rich and 9-rich sequences to form four-stranded G-quartet/9-quartet structures of high stability [75, 76]. Batterby et al. synthesized DNA oligonucleotides that would form duplexes containing a mixture of A•I and G•9 base pairs (Figures 1.6E, 1.6.F) [77]. Duplexes of 14 base pairs in length exhibited a melting temperature around 45°C. For comparison, a DNA duplex with an analogous sequence of A•T and G•C base pairs would have a melting temperature of approximately 60°C under similar experimental conditions [78]. Soon after, Switzer and coworkers reported on the stability of homopurine DNA duplexes in which all base pairs are held together by three hydrogen bonds [79]. Specifically, duplexes with G•9 and D•^{C7}X base pairs, where D is again 2,6-diaminopurine (**13**), but 7-deazaxanthine, or ^{C7}X (i.e., **25** with N7 replaced by CH), was used in place of X due to the tendency for xanthine nucleosides to depurinate. Switzer and coworkers found that duplexes formed with these base pairs are of similar stability as DNA duplexes with A•T and G•C base pairs.

The greater stacking of purine nucleotides inspired our laboratory to investigate if homo-purine oligonucleotides have a greater propensity to undergo non-enzymatic

polymerization than mixed purine-pyrimidine oligonucleotides. Using a system of tetranucleotides that can form mini-helices, it was found that homo-purine strands undergo chemical ligation at least 100-fold more efficiently compared to tetranucleotides with mixed purine/pyrimidine base sequences [80]. These results are consistent with the theory that the larger area of a purine-purine base pair provides a substantially greater free energy for assembly than a purine-pyrimidine base pair in aqueous solution. Since larger surface areas appear to assist in both assembly and non-enzymatic oligomerization reactions one might think that heterocycles even larger than purines, such as nucleotides containing three or more rings (i.e., flavins), would be promising proto-nucleobase candidates. However, neutral aromatics containing more than two rings are very insoluble, leading to the problem of limited availability mentioned above.

1.3.3 Alternative Pyrimidine-Pyrimidine Base Pairs

Motivated by the idea that proto-RNA may have had an achiral backbone [81], Schwartz hypothesized that a pre-RNA could have been synthesized from pyrimidine nucleobases with two pendent alcohol groups [82]. In particular, Schwartz proposed that prebiotic reactions could have produced barbituric acid (BA, **79**) and related pyrimidines in which one or more of the exocyclic oxygens of BA are replaced by an NH_2 group (i.e., **58**, **61**, **67**, **70**, **76**). The nucleic acid backbone proposed in this model would be connected through the C5 position of the pyrimidines (CH_2 in neutral BA), thereby using the base itself as part of the backbone. Schwartz and coworkers synthesized activated versions of two such monomers, but template-directed synthesis reactions with canonical

nucleic acid oligonucleotides as template strands resulted in low yields, which was attributed to the poor stacking of the pyrimidine bases [83].

Expanding on the proposal that peptide nucleic acids could have pre-dated RNA in early life [84], Eschenmoser, Krishnamurthy and coworkers explored the stability of base pairs formed by pyrimidine nucleobases connected to peptide backbones [85]. These researchers realized that four pyrimidines, **41**, **44**, **50**, and **53**, provide a complete set of Watson-Crick pairing faces, as well as an amino group at the C5 positions of sufficient reactivity to allow for coupling to a carboxylic acid group (i.e., amide bond formation). Specifically, these bases were attached to the side chains of glutamic acid residues to create a novel peptide nucleic acid. These peptide-heterocycle conjugates formed hetero-duplexes with RNA and DNA oligonucleotides containing complimentary nucleobases, but demonstrated either a very weak or a non-existent propensity to form homo-duplexes. These studies also brought to light the importance of considering the pK_as of pairing bases when proposing possible proto-RNA base pairs. Moreover, this study led these researchers to conclude that the structure of the recognition elements (of proto-RNA) may have been more critical than the structure of the backbone when it came to nature's "choice" of the first genetic systems [85].

1.3.4 Triazine Base Pairing

The chemical landscape should be expanded beyond the purines and pyrimidines so as not to exclude the possibility that early informational polymers used other heterocycles. As mentioned above, triazines are a class of heterocycles that are closely related to the pyrimidines and known to be formed in the same model prebiotic reactions

that produce pyrimidines and purines [18]. Siegel and Tor have previously proposed that these heterocycles may have functioned in early informational polymers, linking to ribose at the exocyclic amines of amino-substituted triazines [48]. A number of studies have already demonstrated that triazines can function in the place of pyrimidines in the formation of pairing structures. Eschenmoser and coworkers have shown that peptide nucleic acid backbones tagged with complementary triazines can, albeit weakly, form homo-duplexes and cross-pair to form more thermally stable hetero-duplexes with both DNA and RNA [86]. The Bong laboratory has also shown that peptide nucleic acids tagged with melamine (the triazine analog of TAP, **85**) are able to form triplex assemblies with either poly-U or poly–dT [49], demonstrating melamine's compatibility with the extant nucleobases. Additionally, Bong and coworkers have demonstrated that tethered trivalent melamine and cyanuric acid (**91**) molecules can co-assemble in water to form hexameric rosette based assemblies (Figure 1.7) due to the heterocycles ability to pair on two of their edges [87].

1.4 From Molecular Self-Assembly to Proto-RNA and Beyond

1.4.1 Monomeric Assembly through Watson-Crick-Like Pairing in Water

While the iconic Watson-Crick base pairs have long directed the search for pre-RNA towards diad-based informational systems, that the poor tendency of the extant nucleobases to assemble at the monomer level, and the resultant *Paradox of Base Pairing*, might only be resolved by considering heterocycles with the ability to form larger assemblies. This consideration has been supported by both theoretical and

experimental studies which have indicated that the hydrophobic effect is able to drive self-assembly when a hydrophobic surface has an area of 1 nm^2 or greater (which is larger than the diad based surfaces discussed above) [88-90].

For over 50 years it has been known that guanosine and guanosine derivatives (e.g., guanosine mononucleotides, GMP) can self-assemble in water [91], indicating that some heterocycles may potentially be capable of aqueous assembly as monomers. G-tetrads, ion coordinated quartets of guanine, stack in water to form linear noncovalent assemblies [92]. The assembly of these monomers into linear fibers by the stacking of surface areas greater than 1 nm^2 is consistent with the need to form large enough surface areas to drive assembly through the hydrophobic effect. Guanosine at high concentrations will form a meta-stable hydrogel, indicating the presence of long noncovalent assemblies that are noncovalently cross-linked to form a gel matrix. The short gel lifetime of guanosine-based gels are dramatically extended if GMP or a modified guanosine is used instead of guanosine or in combination with guanosine [93-95], potentially due to the lack in favorable packing interactions and, for GMP, charge repulsion between the linear assemblies. Interestingly, when the pH of a GMP solution is raised above pH 6, gelation is lost while the assemblies persist. This may result from increased charge repulsion between assemblies due to the dianionic state of GMP above pH 6, as well as truncation of the supramolecular polymers that is known to occur with an increase in peripheral charge [96]. While guanosine assemblies demonstrate the possibility of nucleobase assembly in water, it is difficult to see how these self-pairing structures could have led to informational polymers given that only one recognition element is utilized.

Molecular self-assembly of non-coupled heterocycles through base-pairing-like interactions has also been demonstrated by Fenniri and coworker by utilizing self-complementary Janus-type bicyclic and tricyclic heterocycles that contain a guanine-like and cytosine-like edge [97, 98]. Similar to the trivalent melamine (**85**) and cyanuric acid (**91**) system reported by Bong, these heterocycles assemble into hexameric rosettes, however, unlike the previous reported rosettes these recognition elements were not tethered. The rosettes also possess a large hydrophobic surface and consequently produce very long supramolecular polymers containing hundreds to thousands of ordered heterocycles. In the initial study, Fenniri et. al. attempted to assemble modified barbituric acid and triaminopyrimidine, however, they did not find evidence of any interaction between these complementary heterocycles [97].

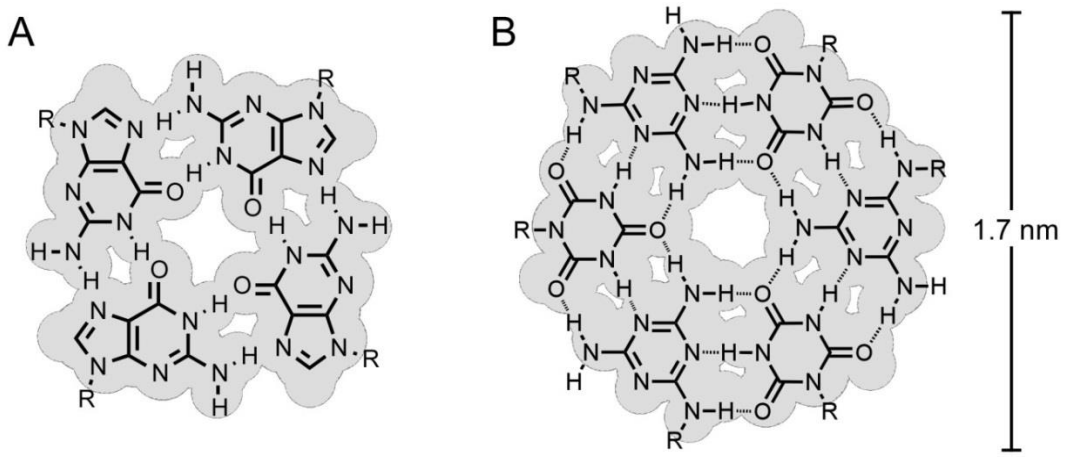


Figure 1.7 Chemical structures of noncovalent macrocycles with space filling representation (grey). (A) G-tetrad assembly motif and (B) melamine-cyanuric acid rosette assembly motif reported to form by Bong and coworkers after tethering. R group on guanine indicates attachment of ribose or ribose derivative and R groups on melamine and cyanuric acid indicate the positions of tethering by a hydrophilic linker.

1.4.2 Molecular Midwife Hypothesis

Alternatively, if informational heterocycles do not properly assemble in water (e.g., the canonical nucleobases) through Watson-Crick-like pairing other molecules may have aided in their assembly and organization, without being covalently incorporated themselves. This proposal by Hud and Anet is termed the "Molecular Midwife Hypothesis" and is supported by the observation small aromatic molecules commonly used as nucleic acid intercalators can enable the assembly and polymerization of short or otherwise non-pairing oligonucleotides (Figure 1.8) [34, 35]. Strand cyclization is problematic for the polymerization of bifunctionalized nucleic acid polymers due to the more entropically favored intramolecular ligation. Prior work in the Hud laboratory has demonstrated the utility of intercalators to concatermerize (tile) activated tetranucleotides at micromolar concentrations by taking advantage of the assembling power of the small molecule and the increased persistence length of the intercalated double strand [99].

Most known intercalators share several chemical traits, including a multi-ring planar surface that inserts between base-pairs, and a cationic charge that aids in solubilizing the heterocycles, molecular recognition with the anionic phosphodiester backbone and potential cation-pi interactions. Binding of these small molecules to nucleic acids has also been shown to strictly obey the nearest-neighbor exclusion principle (NNEP), the negative cooperativity observed when molecules intercalate duplex nucleic acids leading to the binding between every other base pair. Where intercalators markedly differ is in their ability to bind different nucleic acid structures (e.g., duplex, triplex or quadruplex), which has been correlated with the match on an intercalator's surface to that of the base-pair, base-triplet, or G-tetrad [99-102]. However, an additional consideration

for intercalative binding of nucleic acids comes from steric contributions of the sugar phosphate backbone that must go through a conformational change in order to accommodate the intercalator. The energetic contribution of the backbone to the overall free energy of intercalative binding is poorly understood, but has been used to explain the NNEP and other observations [103].

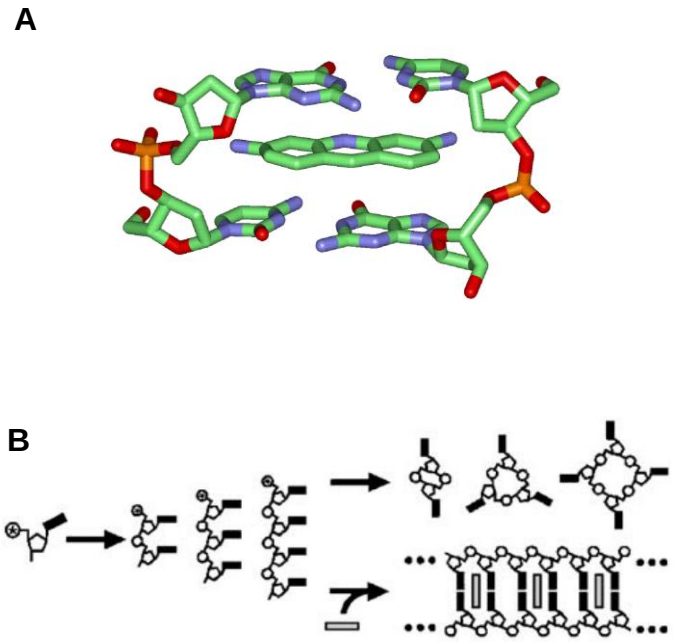


Figure 1.8 Nucleic Acid intercalation example and suppression of strand cyclization. (A) Proflavine inserted between a short C-G dinucleotide duplex. (B) Schematic illistration of the strand cyclizatino problem: as activated monomers are coupled to oligomers cyclization eventually becomes most favorable, however, intercaltion can be used to promote polymers growth. Copywrite, National Academy of Sciences.

1.4.3 Connecting the Pieces: What Originally Linked the First Nucleosides?

Supramolecular assemblies may have selected pairing bases out of complex environments, through Watson-Crick-like interactions and/or small molecule mediated assembly, however, covalent linkages between the monomers would next be required to form true informational polymers.

‘‘Why Nature Chose Phosphates’’ by Westheimer provides an elegant analysis of the suitability of phosphate for its various roles in extant biology [104]. The phosphate group links the nucleotide monomers of contemporary nucleic acids, providing solubility in water and resistance to spontaneous hydrolysis while being readily hydrolyzed by enzymes; nucleic acids utilizing phosphate are thermodynamically unstable but kinetically stable. The positive free energy of formation associated with the phosphodiester linkage in water necessitates the input of energy for polymerization, which is one reason to consider phosphate as a latter addition to nucleic acids. In model prebiotic reactions, chemists often activate phosphates of nucleotides by transient modification with high-energy chemical agents, such as carbodiimide and cyanide-containing compounds. While such reagents are common and useful in synthetic organic chemistry, their prebiotic relevance is questionable [105]. Furthermore, the rate of hydrolysis of an activated phosphate group is comparable to the rate of uncatalyzed phosphodiester bond formation, which can result in inefficient polymerization. It is reasonable to propose that both proto- and pre-RNAs would have benefitted from backbone linkages that were thermodynamically favored (in at least some accessible environment) with low kinetic barriers to hydrolysis—properties that would facilitate activation-free polymerization and would allow environmental switching between

conditions that alternately promote polymerization and hydrolysis. Facile thermodynamically driven switching of this type would promote rapid recycling of materials between polymers and monomers, a property that could have greatly accelerated sequence evolution and emergence of enzymatic activity [106]. Linkages that do not require high energy input would have also been more compatible with supramolecular assemblies that are thermally unstable. Furthermore, low-energy, reversible backbone linkages, for example Schiff-bases [107], allow for selection of thermodynamically favored products, thereby providing an enzyme-free mechanism for accurate templating, error correction, and overcoming kinetic traps such as strand cyclization [35].

1.4.4 Evolving Towards the Canonical Nucleic Acids

It is generally assumed that DNA evolved from RNA in response to the selective pressure to store genetic information in a polymer that is chemically more stable than RNA. However, there is no *a priori* reason to not consider the possibility that the rise of RNA was paralleled by a rise of DNA, or that RNA and DNA are the descendants of a proto-RNA and a proto-DNA, or of a common polymer that diverged into both (Figure 1.9). The coexistence of a genetic polymer and a catalytic polymer that is coded for by the genetic polymer may have been both likely and beneficial [108].

With the synthesis of a theoretical deoxynucleoside intermediate, Szostak and Powner have proposed that deoxyribonucleotides may be prebiotic, and that the first nucleic acid could have consisted of both ribonucleotides and deoxyribonucleotides, termed a mosaic nucleic acid (MNA) [108]. Similar to the mixed 2',5'/3',5' linked RNA

system discussed above, for this scenario some MNA within the same sequence population would be superior for folding into catalytic structures (higher RNA content) while others would serve as better templates and be less prone to hydrolysis (higher DNA content). Overtime homogeneous DNA and RNA polymers would have emerged from the pool of MNA due to selection pressures [108, 109]. In support of the functionality of such a heterogeneous nucleic acid system, *in vitro* selection experiments from libraries containing sequences with randomly incorporated DNA and RNA nucleotides have been effectively used to select for mixed nucleic acid aptamers [110]. As polymer heterogeneity occurs even when starting with a pure solution of activated ribonucleotides (i.e., formation of 2',5'/3',5' linked RNA), future studies involving nucleic acids consisting of more than one sugar type, which may more accurately reflect polymers produced in prebiotic chemical environments, should provide additional insights into the possible origins of RNA, and biopolymers in general.

Another model can be envisioned for the emergence of DNA where two distinct polymers co-exist, similar to the contemporary DNA/RNA system, one being more stable and the other being catalytically superior yet more labile. The prebiotic process that generated the earliest RNA-like polymers could have produced and replicated pre-RNA polymers of random sequences, form which specific sequences would have been selected on a functional basis (e.g., folding, molecular recognition, catalysis) and accumulated within the environment [106]. Subsequently, chemical and/or enzymatic activity (e.g., proto-ribozymes) could have altered the chemical structure of pre-RNA polymers with some transformations leading to a more stable polymer type. Therefore, more abundant and long-lived pre-RNAs ‘aged’ into more stable pre-DNA polymers, while still

maintaining sequence information and templating ability. As one hypothetical mechanism, transition metal catalysis could have modified functional pre-RNAs into more stable polymers, or pre-DNAs, at the polymer level. The appearance of pre-RNA polymers that were capable of converting pre-RNA into pre-DNA could have been aided by iron-based redox chemistry, which has been recently shown to allow RNA to catalyze electron transfer reactions [111]. Regardless of the chemical mechanism that could have converted a pre-RNA into a pre-DNA, the early emergence of a stable genetic molecule would have been highly beneficial to a primitive replication system and set the path for the eventual evolution of ribozymes and/or enzymes of necessary sophistication for the mutually exclusive polymerization of ribonucleotides and dexoyribonucleotides.

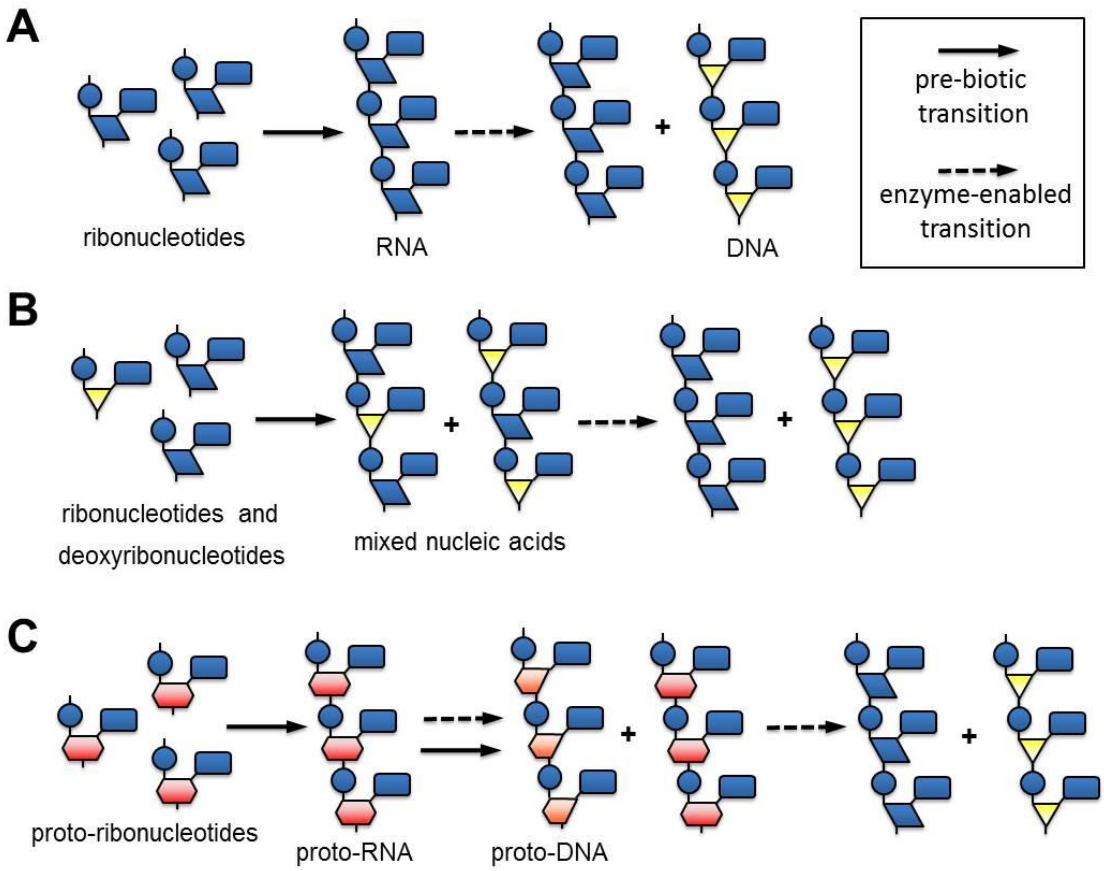


Figure 1.9 Three models for the emergence of DNA. (A) The classic RNA World hypothesis model. There first exists a pool of ribonucleotides. These become polymerized into RNA, DNA forms from biotic processes and eventually takes over the genetic system. (B) Model proposed by Szostak and coworkers in which there first exists a mixed pool of ribonucleotides and deoxyribonucleotides. These nucleotides become stochastically polymerized into mosaic nucleic acids (MNA) and selection pressures eventually lead to homo-RNA and DNA. (C) The Co-RNA-DNA evolution model, in which nucleic acids begin with proto-nucleotides. These proto-nucleotides become polymerized into molecules from which proto-RNA enzymes are selected. By an environmental process (i.e., UV, transition metal catalysis, etc.) or proto-RNA enzyme some of the proto-RNA polymers are converted to proto-DNA. Over time the chemical structure of proto-RNA is refined under evolutionary pressures that act to simultaneously optimize the functionality of successive generations of pre-RNA and pre-DNA, until RNA and DNA are obtained. Solid arrows represent prebiotic transitions and dashed arrows represent transitions that require enzymes. Reproduced with permission from reference 23.

1.5 Objectives

In this dissertation I will describe my work investigating potential molecular building blocks of the first or early nucleic acids. Underscoring these investigations is the hypothesis that the original nucleic acids were more amenable to water-based synthesis, while still chemically similar to the extant nucleobases. By identifying potential barriers for the prebiotic synthesis of informational polymers, we can then propose and test alternative chemistries that may enable a system to overcome the problems associated with *de novo* polymer formation. For example, the spontaneous formation of an RNA-like polymer may be a robust process if the original nucleobases were capable of, 1) self-assembly in water and 2) facile condensation with sugars to form nucleosides in high yield. Through a systematic evaluation of heterocycles that possess physiochemical properties that satisfy both considerations, it may be possible to constrain the chemical space from which the original nucleic acids emerged and to propose likely proto-nucleobase candidates. Additionally, I will address problems related to the inclusion of the ribo-phosphate RNA backbone in an early informational polymer, by investigating oligomers with modified backbone architectures. To understand how matter can obtain life-like qualities through abiotic processes, it is certain that a deep appreciation of the forces that govern this transition is required. For this reason, and beyond the appeal of identifying likely proto-nucleobases, this work also has the potential to provide general insight into molecular self-assembly in aqueous solutions, which has implications in materials development and basic research.

1.6 References

1. Avery, O., C. MacLeod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types.* J Expt Med, 1944. **79**: p. 137-157.

2. Hershey, A.D. and M. Chase, *Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage.* J Gen Phys, 1952. **36**: p. 39-56.

3. Watson, J.D. and F.H.C. Crick, *Genetical implications of the structure of deoxyribose nucleic acid.* Nature, 1953. **171**: p. 964-967.

4. Crick, F., *Central Dogma of Molecular Biology.* Nature, 1970. **226**: p. 561-563.

5. Gesteland, R.F., T.R. Cech, and J.F. Atkins, *The RNA World*. 3rd ed. 2006, Cold Spring harbor, NY: Cold Spring Harbor Laboratory Press. 768.

6. Kruger, K., et al., *Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell, 1982. **31**: p. 147-157.

7. Guerrier-Takada, C., et al., *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.* Cell, 1983. **35**: p. 849-857.

8. Carter, A., et al., *Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics.* Nature, 2000. **407**: p. 340-348.

9. Hsiao, C., et al., *Peeling the onion: ribosomes are ancient molecular fossils.* Mol Biol Evol, 2009. **26**: p. 2415-2425.

10. Noller, H.F., V. Hoffarth, and L. Zimniak, *Unusual resistance of peptidyl transferase to protein extraction procedures.* Science, 1992. **256**: p. 1416-1419.

11. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 Ångstrom resolution.* Science, 2000. **289**: p. 905-920.

12. Oró, J., *Synthesis of adenine from ammonium cyanide.* Biochem Biophys Res Comm., 1960. **2**: p. 407-412.

13. Decker, P., P. Schweer, and R. Pohlmann, *Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectroscopy of n-butoxime trifluoroacetates on OV-225.* J. Chromatogr, 1982. **244**: p. 281-291.

14. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis. VI. synthesis of purine nucleosides.* J Mol Biol, 1972. **67**: p. 25-33.

15. Orgel, L.E., *Prebiotic chemistry and the origin of the RNA world.* Crit Rev Biochem Mol Biol, 2004. **39**: p. 99-123.

16. Benner, S.A., H.-J. Kim, and M.A. Carrigan, *Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA.* Acc Chem Res, 2012. **45**: p. 2025-2034.

17. Saladino, R., et al., *Formamide chemistry and the origin of informational polymers.* Chem Biodiv, 2007. **4**: p. 694-720.

18. Menor-Salvan, C., et al., *Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases.* Chem Eur J, 2009. **15**: p. 4411-4418.

19. Larralde, R., M.P. Robertson, and S.L. Miller, *Rates of decomposition of ribose and other sugars: implications for chemical evolution.* Proc Natl Acad Sci U S A, 1995. **92**: p. 8158-8160.

20. Kim, H.J., et al., *Synthesis of carbohydrates in mineral-guided prebiotic cycles.* J Am Chem Soc, 2011. **133**: p. 9457-9468.

21. Grew, E.S., J.L. Bada, and R.M. Hazen, *Borate minerals and origin of the RNA world.* Orig Life Evol Biosph, 2011. **41**: p. 307-316.

22. Hud, N.V., et al., *The origin of RNA and 'My Grandfather's Axe'.* Chem Biol, 2013. **20**: p. 466-474.

23. Cafferty, B.J. and N.V. Hud, *Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology.* Curr Opin Chem Biol, 2014. **22**: p. 146-157.

24. Rios, A.C., H.T. Yu, and Y. Tor, *Hydrolytic fitness of N-glycosyl bonds: comparing the deglycosylation kinetics of modified, alternative, and native nucleosides.* J Phys Org Chem, 2014. **28**: p. 173-180.

25. Rios, A.C. and Y. Tor, *On the Origin of the Canonical Nucleobases: An Assessment of Selection Pressures across Chemical and Early Biological Evolution.* Isr J Chem, 2013. **53**: p. 469-483.

26. Weber, A.L., *Prebiotic sugar synthesis: hexose and hydroxy acid synthesis from glyceraldehyde catalyzed by iron(III) hydroxide oxide.* J Mol Evol, 1992. **35**: p. 1-6.

27. Sagi, V.N., et al., *Exploratory experiments on the chemistry of the "glyoxylate scenario": Formation of ketosugars from dihydroxyfumarate.* J Am Chem Soc, 2012. **134**: p. 3577-3589.

28. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis: VII. solid-state synthesis of purine nucleosides.* J Mol Evol, 1972. **1**: p. 249-257.

29. Sanchez, R.A. and L.E. Orgel, *Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides.* J Mol Biol, 1970. **47**: p. 531-543.

30. Powner, M.W., B. Gerland, and J.D. Sutherland, *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions.* Nature, 2009. **459**: p. 239-242.

31. Ts'o, P., I. Melvin, and A. Olson, *Interaction and association of bases and nucleosides in aqueous solutions.* J Am Chem Soc, 1963. **85**: p. 1289-1296.

32. Hud, N.V. and F.A.L. Anet, *Intercalation-mediated synthesis and replication: a new approach to the origin of life.* J Theor Biol, 2000. **205**: p. 543-562.

33. Engelhart, A.E. and N.V. Hud, *Primitive genetic polymers.* Cold Spring Harbor Perspectives in Biology, 2010. DOI:10.1101/cshperspect.a002196.

34. Jain, S.S., et al., *Enzymatic behavior by intercalating molecules in a template-directed ligation reaction.* Angew Chem Int Ed, 2004. **43**: p. 2004-2008.

35. Hud, N.V., et al., *Addressing the problems of base pairing and strand cyclization in template-directed synthesis.* Chem Biodivers, 2007. **4**: p. 768-783.

36. Eschenmoser, A., *Chemical etiology of nucleic acid structure.* Science, 1999. **284**(5423): p. 2118-2124.

37. Orgel, L.E., *Evolution of the genetic apparatus.* J Mol Biol, 1968. **38**: p. 381-393.

38. Orgel, L.E., *The origin of life - a review of facts and speculations.* TIBS, 1998. **23**: p. 491-495.

39. Switzer, C., S.E. Moroney, and S.A. Benner, *Enzymatic incorporation of a new base pair into DNA and RNA.* J Am Chem Soc, 1989. **111**: p. 8322-8323.

40. Piccirilli, J.A., et al., *Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet.* Nature, 1990. **343**: p. 33-37.

41. Geyer, C.R., T.R. Battersby, and S.A. Benner, *Nucleobase pairing in Watson-Crick-like genetic expanded information systems.* Structure, 2003. **11**: p. 1485-1498.

42. Pinheiro, V.B., et al., *Synthetic genetic polymers capable of heredity and evolution.* Science, 2012. **336**(6079): p. 341-344.

43. Sefah, K., et al., *In vitro selection with artificial expanded genetic information systems.* Proc Natl Acad Sci U S A, 2014. **111**: p. 1449-1454.

44. Taylor, A.I., et al., *Catalysts from synthetic genetic polymers.* Nature, 2015. **518**: p. 427-430.

45. Zhang, S.L., et al., *Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer.* Proc Natl Acad Sci U S A, 2013. **110**: p. 17732-17737.

46. Hayatsu, R., et al., *Origin of organic matter in early solar system 2. Nitrogen compounds.* Geochim Cosmochim Acta, 1968. **32**: p. 175-190.

47. Hayatsu, R., et al., *Purines and triazines in murchison meteorite.* Geochim Cosmochim Acta, 1975. **39**: p. 471-488.

48. Hysell, M., J.S. Siegel, and Y. Tor, *Synthesis and stability of exocyclic triazine nucleosides.* Org Boimol Chem, 2005. **3**: p. 2946-2952.

49. Xia, X., X. Piao, and D. Bong, *Bifacial Peptide Nucleic Acid as an Allosteric Switch for Aptamer and Ribozyme Function.* J Am Chem Soc, 2014. **136**: p. 7265-7268.

50. Callahan, M.P., et al., *Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases.* Proc Natl Acad Sci U S A, 2011. **108**: p. 13995-13998.

51. Sanchez, R.A., J.P. Ferris, and L.E. Orgel, *Studiesin prebiotic synthesis.4. Conversoin of 4-aminoimidazole-5-carbonitrile derivatives to purines.* J Mol Biol, 1968. **38**: p. 121-128.

52. Miyakawa, S., H.J. Cleaves, and S.L. Miller, *The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions.* Orig Life Evol Biosph, 2002. **32**: p. 209-218.

53. Hartel, C. and M.W. Gobel, *Substitution of adenine by purine-2,6-diamine improves the nonenzymatic oligomerization of ribonucleotides on templates containing thymidine.* Helvetica Chimica Acta, 2000. **83**: p. 2541-2549.

54. Albert, A. and D.J. Brown, *Purine studies. Part 1. Stability to acid and alkali - solubility - ionization - comparison with pteridines.* Journal of the Chemical Society, 1954(JUN): p. 2060-2071.

55. Bean, H.D., et al., *Formation of a β-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction.* J Am Chem Soc, 2007. **129**: p. 9556-9557

56. Gildea, B. and L.W. McLaughlin, *The synthesis of 2-pyrimidinone nucleosides and their incorporation into oligodeoxynucleotides.* Nucleic Acids Res, 1989. **17**: p. 2261-2281.

57. Sheng, Y., et al., *A comprehensive investigation of the energetics of pyrimidine nucleoside formation in a model prebiotic reaction.* J Am Chem Soc, 2009. **131**: p. 16088-16095.

58. Jang, Y.H., et al., *First principles calculations of the tautomers and pK_a values of 8-oxoguanine: Implications for mutagenicity and repair.* Chemical Research in Toxicology, 2002. **15**: p. 1023-1035.

59. Engelhart, A.E., T.H. Morton, and N.V. Hud, *Evidence of strong hydrogen bonding by 8-amino-guanine.* Chem Commun, 2009: p. 647-649.

60. Joyce, G.F. and L. Orgel, *Progress toward understanding the origin of the RNA world*, in *The RNA World*, R. Gesteland, T.R. Cech, and R. Atkin, Editors. 2006, CSHL Press: Cold Spring Harbor. p. 23-56.

61. Stribling, R. and S.L. Miller, *Attempted nonenzymatic template-directed oligomerizations on a polyadenylic-acid template - implications for the nature of the 1st genetic material.* J Mol Evol, 1991. **32**: p. 282-288.

62. Hill, A.J., L. Orgel, and T. Wu, *The limits of template-directed synthesis with nucleoside-5'-phosphoro(2-methyl)imidazolides.* Orig Life Evol Biosph, 1993. **23**: p. 285-290.

63. Crick, F.H.C., *The origin of the genetic code.* J. Mol. Biol., 1968. **38**: p. 367-379.

64. Hill, A. and L.E. Orgel, *Synthesis of adenine from HCN tetramer and ammonium formate.* Orig Life Evol Biosph, 2002. **32**: p. 99-102.

65. Orgel, L.E., *Prebiotic Adenine Revisited: Eutectics and Photochemistry.* Orig Life Evol Biosph, 2004. **34**: p. 361-369.

66. Saladino, R., et al., *A possible prebiotic synthesis of purine, adenine, cytosine, and 4(3H)-pyrimidinone from formamide: Implications for the origin of life.* Bioorg Med Chem, 2001. **9**: p. 1249-1253.

67. Hudson, J.S., et al., *A unified mechanism for abiotic adenine and purine synthesis in formamide.* Angew Chem Int Ed, 2012. **51**: p. 5134-5137.

68. Barks, H.L., et al., *Guanine, adenine, and hypoxanthine production in UV-irradiated formamide solutions: Relaxation of the requirements for prebiotic purine nucleobase formation.* ChemBioChem, 2010. **11**: p. 1240-1243.

69. Zubay, G. and T. Mui, *Prebiotic synthesis of nucleosides.* Origins Life Evol. Biosph., 2001. **31**: p. 87-102.

70. Augustyns, K., et al., *Influence of the incorporation of 1-(2,3-dideoxy-beta-D-erythro-hexopyranosyl)thymine on the enzymatic stability and base-pairing properties of oligodeoxynucleotides.* B Soc Chim Belg, 1992. **101**: p. 119-130.

71. Albert Eschenmoser, M.D., *Warum Pentose- und nicht Hexose-Nucleinsauren?? Teil I. Einleitung und Problemstellung, Konformationsanalyse fur Oligonucleotid-Ketten aus 2',3'-Dideoxyglucopyranosyl-Bausteinen ('Homo-DNS') sowie Betrachtungen zur Konformation von A- und B-DNS.* Helv Chim Acta, 1992. **75**: p. 218-259.

72. Groebke, K., et al., *Why pentose- and not hexose-nucleic acids? Purine-purine pairing in homo-DNA: guanine, isoguanine, 2,6-diaminopurine, and xanthine.* Helv Chim Acta, 1998. **81**: p. 375-474.

73. SantaLucia, J. and D. Hicks, *The thermodynamics of DNA structural motifs.* Annu Rev Bioph Biomol Struct, 2004. **33**: p. 415-440.

74. Eschenmoser, A., *Etiology of potentially primordial biomolecular structures: From vitamin B12 to the nucleic acids and an inquiry into the chemistry of life's origin: A retrospective.* Angew Chem Int Ed, 2011. **50**: p. 12412-12472.

75. Howard, F.B. and H.T. Miles, *Interaction of poly(A) and poly(I), a reinvestigation.* Biochemistry, 1977. **16**: p. 4647-4650.

76. Roberts, C., J.C. Chaput, and C. Switzer, *Beyond guanine quartets: cation-induced formation of homogenous and chimeric DNA tetraplexes incorporating iso-guanine and guanine.* Chem Biol, 1997. **4**: p. 899-908.

77. Battersby, T.R., M. Albalos, and M.J. Friesenhahn, *An unusual mode of DNA duplex association: Watson-Crick interaction of all-purine deoxyribonucleic acids.* Chem Biol, 2007. **14**: p. 525-531.

78. SantaLucia, J., N. Peyret, and P. Saro, *HyTher: Prediction of Nucleic Acid Hybridization Thermodynamics*. 1998, Chemistry Department, Wayne State University.

79. Heuberger, B.D. and C. Switzer, *An alternative nucleobase code: characterization of purine-purine DNA double helices bearing guanine-isoguanine and diaminopurine-7-deaza-xanthine base pairs.* ChemBioChem, 2008. **9**: p. 2779-2783.

80. Kuruvilla, E., G.B. Schuster, and N.V. Hud, *Enhanced non-enzymatic ligation of homo-purine miniduplexes: Support for greater base stacking in a pre-RNA world.* ChemBioChem, 2013: p. 10.1002/cbic.201200601.

81. Joyce, G.F., et al., *The case for an ancestral genetic system involving simple analogs of the nucleotides.* Proc Natl Acad Sci U S A, 1987. **84**: p. 4398-4402.

82. Schwartz, A.W., *Nucleotide analogs based on pentaerythritol - an hypothesis.* Orig Life Evol Biosph, 1993. **23**: p. 185-194.

83. van Vliet, M.J., J. Visscher, and A.W. Schwartz, *Hydrogen-bonding in the template-directed oligomerization of a pyrimidine nucleotide analog.* J Mol Evol, 1995. **41**: p. 257-261.

84. Nielsen, P.E., *Peptide nucleic acids and the origin of life.* Chem Biodiv, 2007. **4**: p. 1996-2002.

85. Mittapalli, G.K., et al., *Mapping the landscape of potentially primordial informational oligomers: oligodipeptides tagged with 2,4-disubstituted 5-aminopyrimidines as recognition elements.* Angew Chem Int Ed, 2007. **46**: p. 2478-2484.

86. Mittapalli, G.K., et al., *Mapping the landscape of potentially primordial informational oligomers: oligodipeptides and oligodipeptoids tagged with triazines as recognition elements.* Angew Chem Int Ed, 2007. **46**: p. 2470-2477.

87. Ma, M. and D. Bong, *Determinants of cyanuric acid and melamine assembly in water.* Langmuir, 2011. **27**: p. 8841-53.

88. Davis, J.G., et al., *Water structural transformation at molecular hydrophobic interfaces.* Nature, 2012. **491**: p. 582-585.

89. Chandler, D., *Interfaces and the driving force of hydrophobic assembly.* Nature, 2005. **437**: p. 640-647.

90. Raschke, T.M., J. Tsai, and M. Levitt, *Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solutes in water.* Proc Natl Acad Sci U S A, 2001. **98**: p. 5965-5969.

91. Gellert, M., M.N. Lipsett, and D.R. Davies, *Helix Formation by Guanylic Acid.* Proc Natl Acad Sci U S A, 1962. **48**: p. 2013-2018.

92. Davis, J.T., *G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry.* Angew Chem Int Ed, 2004. **43**: p. 668-698.

93. Buerkle, L.E., et al., *Tailoring the properties of guanosine-based supramolecular hydrogels.* Langmuir, 2009. **25**: p. 8833-8840.

94. Li, Z., et al., *Structure and gelation mechanism of tunable guanosine-based supramolecular hydrogels.* Langmuir, 2010. **26**: p. 10093-10101.

95. Yu, Y., et al., *Tunable thermoassociation of binary guanosine gels.* J Phys Chem B, 2008. **112**: p. 1130-1134.

96. Schaefer, C., et al., *Controlling the Cooperativity in the Supramolecular Polymerization of Ionic Discotic Amphiphiles via Electrostatic Screening.* ACS Macro Letters, 2012. **1**: p. 830-833.

97. Fenniri, H., et al., *Helical rosette nanotubes: Design, self-assembly, and characterization.* J Am Chem Soc, 2001. **123**: p. 3854-3855.

98. Bozsonyi, G., et al., *Water-soluble J-type rosette nanotubes with giant molar ellipticity.* J Am Chem Soc, 2010. **132**: p. 15135-15139.

99. Horowitz, E.D., et al., *Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world.* Proc Natl Acad Sci U S A, 2010. **107**: p. 5288-5293.

100. Çetinkol, Ö.P., et al., *Submicromolar, selective G-quadruplex ligands from one pot: thermodynamic and structural studies of human telomeric DNA binding by azacyanines.* ChemBioChem, 2008. **9**: p. 1889-1892.

101. Persil, Ö. and N.V. Hud, *Harnessing DNA intercalation.* Trends Biotechnol, 2007. **25**: p. 433-436.

102. Buckley, R., et al., *Molecular Recognition of Watson-Crick-Like Purine-Purine Base Pairs.* ChemBioChem, 2011. **12**: p. 2155-2158.

103. Voet, D., *Intercalation complexes of DNA.* Nature, 1977. **269**: p. 285-286.

104. Westheimer, F.H., *Why Nature chose phosphates.* Science, 1987. **235**: p. 1173-1178.

105. Keefe, A.D. and S.L. Miller, *Are polyphosphates or phosphate esters prebiotic reagents?* J Mol Evol, 1995. **41**: p. 693-702.

106. Walker, S.I., M.A. Grover, and N.V. Hud, *Universal sequence replication, reversible polymerization and early functional biopolymers: A model for the initiation of prebiotic sequence evolution.* PLoS ONE, 2012. **7**: p. e34166.

107. Engelhart, A.E., et al., *Nonenzymatic ligation of DNA with a reversible step and a final linkage that can be used in PCR.* ChemBioChem, 2012. **13**: p. 1121-1124.

108. Powner, M.W., S.L. Zheng, and J.W. Szostak, *Multicomponent assembly of proposed DNA precursors in water.* J Am Chem Soc, 2012. **134**: p. 13889-13895.

109. Ritson, D.J. and J.D. Sutherland, *Conversion of biosynthetic precursors of RNA to those of DNA by photoredox chemistry.* J Mol Evol, 2014. **78**: p. 245-250.

110. Trevino, S.G., et al., *Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity.* Proc Natl Acad Sci U S A, 2011. **108**: p. 13492-13497.

111. Hsiao, C., et al., *RNA with iron(II) as a cofactor catalyses electron transfer.* Nature Chem, 2013. **5**: p. 525-528.

112. Cafferty, B.J. and N.V. Hud, *Was a Pyrimidine-Pyrimidine Base Pair the Ancestor of Watson-Crick Base Pairs? Insights from a Systematic Approach to the Origin of RNA.* Isr J Chem, 2015. DOI: 10.1002/ijch.201400206

CHAPTER 2

SUPRAMOLECULAR POLYMERIZATION OF NUCLEOBASE ANALOGS IN WATER[b]

2.1 Introduction

Elucidating the physiochemical principles that govern molecular self-assembly in water is essential for understanding biological systems at the molecular level, for developing new materials, and for nanotechnology [1, 2]. Typically, aqueous assembly results from a combination of hydrogen bonding, dispersion forces, ion pairing, metal-ligand coordination, hydrophobic and stacking interactions [3, 4]. Current efforts to improve the self-assembly of synthetic molecules in water include increasing the number of these interactions and covalently joining multiple recognition units with a flexible linker [5, 6] which promotes assembly as a result of coupling favorable noncovalent interactions. When unidirectional interactions are employed, long linear structures, termed supramolecular polymers, can be formed [7, 8]. Based on the mechanism of their formation, supramolecular polymers can be characterized as either isodesmic or cooperative [8]. An isodesmic assembly mechanism is characterized by a step growth process and large polydispersity. A cooperative assembly mechanism is characterized by the coexistence of monomers and very long noncovalent assemblies above a minimal assembly concentration; similar to micelle formation observed during fatty acid aggregation (i.e., critical micelle concentration) [9].

b. This chapter was adapted from previously published work and is reproduced with permission. Cafferty, BJ; Gallego, I; Chen, MC; Farley, KI; Eritja; Hud, NV "Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues." *J. Am. Chem. Soc.* (2013).

Unique to water based systems, the hydrophobic effect can be the driving force behind cooperative supramolecular polymerization. Hierarchical assembly methods involving intermediate structures with large planar hydrophobic surfaces that stack to form noncovalent polymers have been demonstrated using both synthetic [10-12] and biologically derived molecules [13]. Both theoretical studies and recent experimental results indicate that hydrophobic surfaces must be larger than a minimum size (ca. 1 nm^2) in order for the hydrophobic effect to facilitate molecular self-assembly [14-16]. The ability to rationally employ the hydrophobic effect, in conjunction with other noncovalent interactions, could facilitate the design of molecules capable of most efficient water-based self-assembly including large supramolecular structures with predictable architectures. However, a simple description of what is needed to harness the hydrophobic effect is still lacking, which limits its use as a design parameter for artificial systems.

For over twenty years the triazines melamine (MA) and cyanuric acid (Cy) and the related pyrimidines 2,4,6-triaminopyrimidine (TAP) and barbaturic acid (BA) have been used as recognition units to construct many different three-dimensional noncovalent assemblies, including the formation of six-membered rosettes (Figure 2.1) [17, 18]. Early studies with these motifs emphasized the importance of hydrogen bonding in supramolecular assemblies and were accordingly carried out in “noncompeting” aprotic solvents. Later attempts to assemble the complementary monomeric molecules in water resulted in the formation of irregular precipitates [19] or the total loss of molecular recognition [10], unless multiple units are joined through a hydrophilic covalent linker [6]. We were intrigued by the potential of these historically significant assembly motifs

because the surface area of a six-membered rosette (e.g., TAP_3Cy_3) is 1.7 nm^2, and secondly, the similarity of these recognition units to nucleic acid bases presents the possibility of understanding the energetics of assembled structures in terms of well-characterized molecular interactions. Specifically, decades of research into RNA and DNA structures has produced some of the most accurate numbers available for the standard state free energies of H-bond and stacking interactions in water [20, 21].

With the concept of a minimum hydrophobic surface area in mind, we designed a system that achieves cooperative self-assembly in water. Two weakly interacting low molecular weight monomers (Cy and a modified TAP) are shown to form extremely long supramolecular polymer assemblies that retain water solubility. The observed equilibrium between only free monomers and supramolecular assemblies is in excellent agreement with literature values for the free energy of nucleic acid base interactions, as well as the calculated free energy penalty for the exposure of hydrophobic structures in water. The results of our study have implications for the design of new hydrogel forming molecules, and may also provide insights into the origin of the first RNA-like polymer.

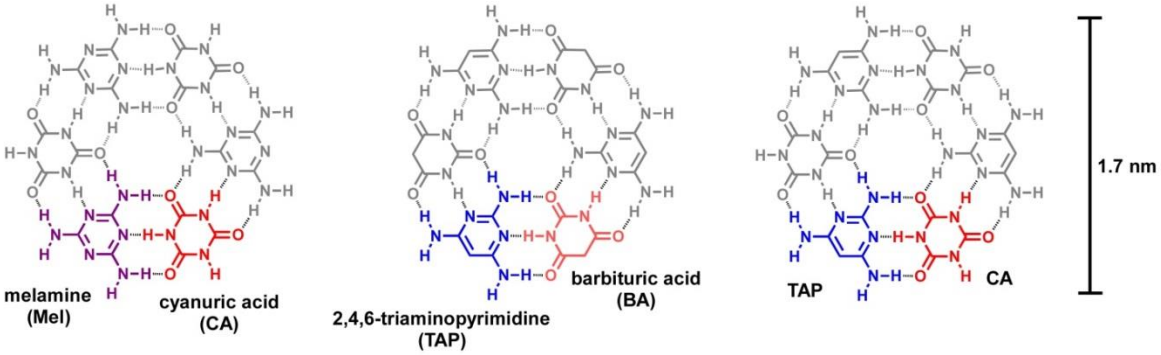


Figure 2.1 Chemical structures and assembly of heterocycles used to form rosette motifs.

2.2 Experimental Procedures

2.2.1 Materials

2,4,6-triaminopyrimidine (TAP) and cyanuric acid (Cy) were purchased from Acros Organic, succinic anhydride was purchased from Alfa Aesar. All materials were used as received.

2.2.2 Synthesis and Characterization

3-(2,6-diaminopyrimidin-4-ylcarbamoyl)propanoic acid, TAPAS. 2,4,6-triaminopyrimidine (0.5 g, 4 mmols) was added to anhydrous *N*,*N*-dimethylformamide (DMF), 30 ml, heated to 60 °C and stirred under argon to dissolve the triaminopyrimidine. The temperature of the solution was then reduced to 40 °C and succinic anhydride was added (0.4 g, 4 mmols) in portions over 30 min. The reaction was stopped after 15 h and the DMF was removed in vacuum (52%, yield). Because TAPAS has the propensity to self-associate in water as a zwitterion at neutral pH, this molecule could be purified from most of the starting material by washing with water. The reaction product was dissolved in alkaline water (final pH 10) and purified using Sephadex DEAE anion exchange media (20 mM tris buffer; NH_4OAc 20 mM elution buffer, pH 8) with the elution fractions monitored by UV-Vis. Product was then washed with water to remove Tris before being resuspended in water (pH 11) and stored at -20 °C

TAPAS purity was determined by NMR and LCMS (Figure 2.2). ^1H NMR was performed in DMSO on a Varian Mercury 400 MHz NMR. LC analysis was performed with an Agilent 1200 Infinity system using a Luna C18(2) column by Phenomenex. The eluting solvent was run isocratic; 5:95, acetonitrile: 5% formic acid solution, over the

course of 15 min. MS analysis was performed using an Agilent 6130 quadruple mass spectrometer run in positive ion mode. Accurate mass analysis was performed using Waters *SYNAPT G2* High Definition MS System. ^{1}H NMR (400 MHz, DMSO): δ 11.14 p.p.m. (s, 1H), 6.49 (s, 1H), 6.11 (s, 2H), 5.62 (s, 2H), 2.34 (t, J=5.7 Hz, 2H), 2.13 (t, J= 5.9 Hz, 2H); ^{13}C NMR (100 MHz, DMSO): δ 174.6, 172.0, 166.4, 163.4, 158.1, 82.9, 31.8, 21.4; UV/vis λ_{max} 288 nm (pH > 7); HRMS (m/z): $[M]^{+}$ calculated for $C_8H_{12}N_5O_3$, 226.0940; found, 226.0950.

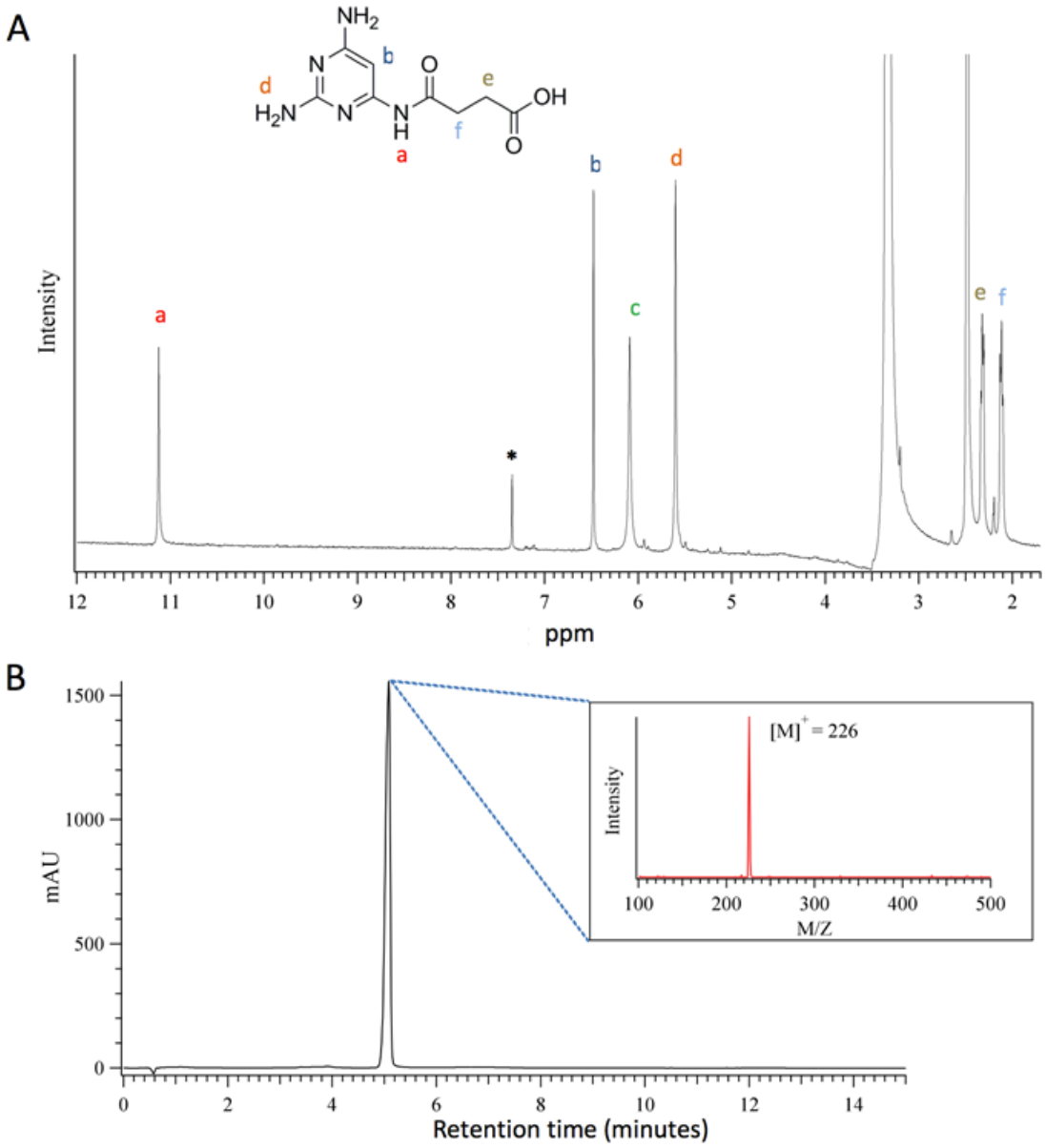


Figure 2.2 ^{1}H NMR and HPLC determination of TAPAS sample purity. (A) NMR spectrum showing all observed ^{1}H resonances of TAPAS. Signal at 7.37 ppm (*****) corresponds to benzene, which was added to the sample as an internal standard to confirm the TAPAS extinction coefficient that was determined by both UV and gravimetric measurements. Signals at 3.33 and 2.50 ppm correspond to H_2O and the residual DMSO peak, respectively. (B) LCMS analysis of purified TAPAS. Chromatograph was generated by monitoring absorbance at 288 nm of HPLC elution. Inset is a mass spectrum of the chromatograph peak at 4.5 min, which reveals a single species with 226 m/z, the mass of protonated TAPAS.

2.2.3 TAPAS pK_a Determination

The pK_a of the aromatic ring of TAPAS was determined by following change in absorbance maxima between protonation and deprotonation of the pyrimidine ring (charged state Abs max at 282 nm; neutral state Abs max at 288 nm). A two component buffer system was used: a solution of 88 mM TAPAS in 200 mM phosphate was titrated with a solution containing 88 mM TAPAS and 100 mM citric acid. At each addition a UV spectrum was acquired and the solution pH was measured with a pH probe. The resulting data was fit with a sigmoidal curve that revealed a pK_a of 6.5. The TAPAS extinction coefficient was determined by alkaline hydrolysis of TAPAS back to triaminopyrimidine and succinate. A solution of 1 mM TAPAS and 1 M NaOH was incubated at 55 °C for 4 h in an airtight container; complete hydrolysis was verified by HPLC. The extinction coefficient of triaminopyrmidine at pH 8.1 was determined to be 10,100 $L \cdot mol^{-1} \cdot cm^{-1}$ by dissolving a known amount of dry sample with a tris-Cl solution to various concentrations and measuring absorbance. The TAPAS extinction coefficient at pH 8.1 (Tris-Cl) was determined to be 6930 $L \cdot mol^{-1} \cdot cm^{-1}$.

2.2.4 Sample Preparation

Stock solutions of TAPAS and Cy each contained 100 mM of either molecule, and were adjusted to pH 10 with NaOH to maintain solubility. TAPAS stock solutions were stored at -20 °C and Cy stock solutions were prepared on the day of use. Assemblies were formed by adding TAPAS to a solution containing an equivalent amount of Cy and either 200 mM sodium phosphate buffer for pH 7 solutions or 200 mM Tris-HCl for pH 8 solutions; the resulting solution was adjusted to the appropriate pH with HCl. Unless

otherwise noted, assembly solutions were made through dilution of a 30 mM concentrated sample with 200 mM sodium phosphate buffer, followed by sonication, vortexing and equilibration at 20 °C before analysis. All assembly solutions were analyzed within 5 h of preparation.

2.2.5 Instrumental Methods

Spectroscopic analysis of assemblies was performed using UV-Vis, ^{1}H NMR and DLS. UV-Vis analysis was performed on an Agilent 8453 spectrophotometer equipped with an 89090A temperature controller. Cells of different path lengths (1, 0.1 and 0.01 mm) were used depending on the concentration of the sample to maintain an optical density below 1.2. ^{1}H NMR spectra were collected on a Bruker DRX-500 and were the sum of 64 transients. All samples contained 10% D_2O and were observed using a WATERGATE pulse sequence. Cacodylic acid was used as an internal standard. DLS was performed using a Protein Solutions DynaPro-99 module with all solutions passed through a 0.22 micron filter prior to measurement.

AFM imaging was performed on freshly cleaved mica that was pre-activated by treatment with 20 mM $MgCl_2$ for 1-2 h. 2-µl of the sample was spin coated for 30 sec at 2500 rpm. AFM imaging was performed with a Nanoscope IIIa (Digital Instruments) in tapping mode in air, using Si tips (Vistaprobes, 40 N/m; MikroMasch, 5.7 N/m). TEM samples were prepared by placing a copper carbon-coated grid (300 mesh; Ted Pella) over a 5-µl drop of the assembly solution for 20 s. The sample was stained by placing the grid two times over a 5-µl drop of 1% uranyl acetate for 20 s. The micrographs were collected in a JEOL 100 CX electron microscope operated at 100 kV.

2.3 Results and Discussion

2.3.1 TAPAS-Cy Assembly Motif Design

A two-component self-assembling system was constructed, consisting of Cy and a TAP that was modified with a succinate group (TAPAS) on one exocyclic amine (Figure 2.3). This modification was performed to sterically inhibit the formation of non-rosette assemblies (e.g., ribbons and sheets) which would otherwise lead to co-precipitation of the unmodified TAP and Cy motifs [22]. For example, mixtures of unmodified TAP and Cy immediately precipitate when mixed in solution at concentrations above 10 mM. Additionally, the charge on the peripheral group enhances the solubility of both the monomer and the large supramolecular structures that are generated upon assembly. While the system contains two distinct monomers containing three equivalent H-bonding faces, only one of the heterocycles must be modified.

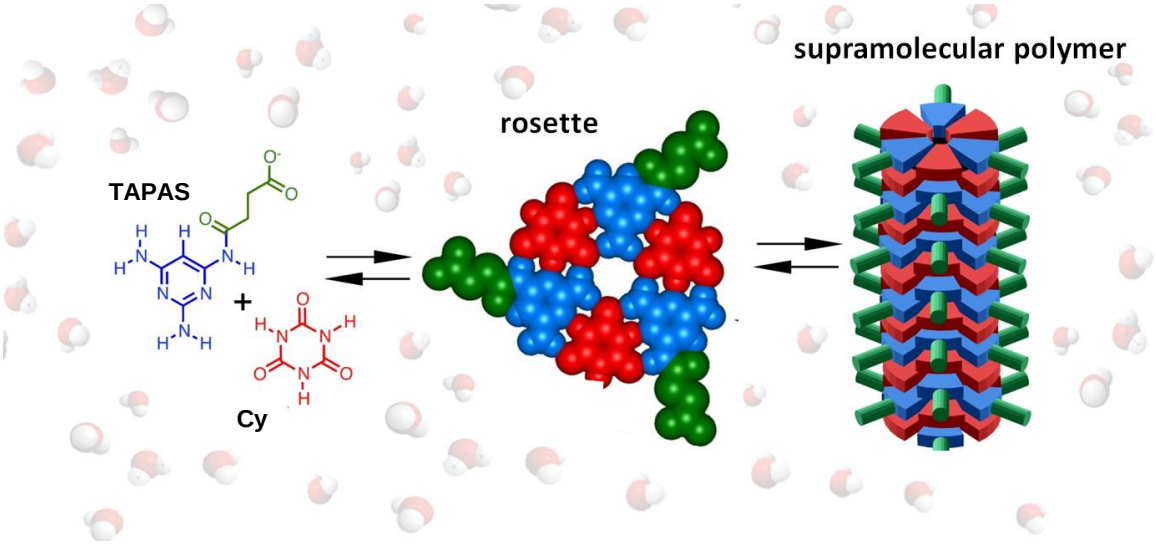


Figure 2.3 Chemical structures of TAPAS (2,4,6-triaminopyrimidine conjugated with succinate) and Cy (cyanuric acid), TAPAS-Cy rosette, and proposed polymer formed by stacked rosettes.

2.3.2 TAPAS and Cy Co-Assemble in Water

When equimolar mixtures of TAPAS and Cy were made by combining both molecules at concentrations above 5 mM a hydrogel formed (initially at pH 10 and then reduced to 7 by titration with HCl or the addition of 200 mM sodium phosphate buffer) (Figure 2.4). The rates of gel formation and gel rigidity increase with monomer concentration. For example, solutions containing each monomer at 5 mM gel within 10 minutes while 30 mM solutions gel in seconds. In unbuffered solutions initially at pH 7 or pH 8, an increase in pH was also observed upon gel.

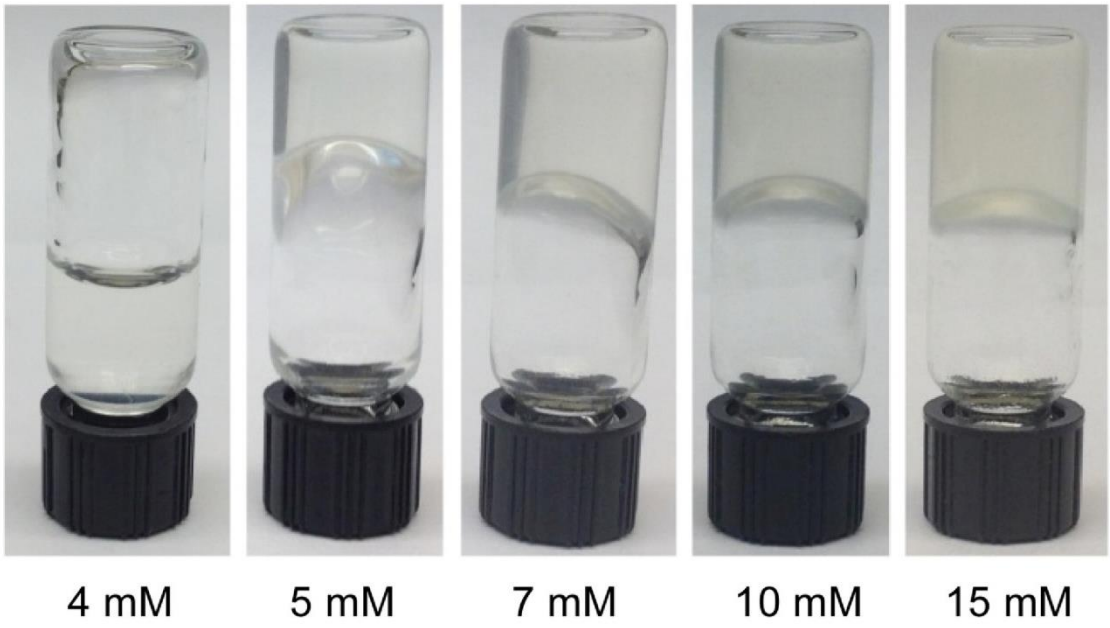


Figure 2.4 Inverted bottle test demonstrating gelation as a function of concentration. Each solution was composed of equimolar TAPAS and Cy , buffered with and 200 mM phosphate (pH 7) and equilibrated at room temperature. Note, the 4 mM sample showed no gelation after at least 2 h.

Gelation of low molecular-weight monomers is a well-known phenomenon that can be used to generate highly dynamic soft-materials that have been demonstrated to function in a wide variety of biomolecular applications [7, 23-25]. The gelation typically results from the formation of long polymers that non-covalently cross-link to produce a three-dimensional matrix that inhibits the bulk flow of water. A related system are the G-tetrads (formed from guanosine, guanosine monophosphate or other guanosine derivatives) that are well known to stack and form supramolecular polymers that cause gelation at sufficiently high concentrations [13, 26].

The architecture of the TAPAS-Cy assemblies was visualized by atomic force microscopy (AFM) in tapping mode and transmission electron microscopy (TEM). Both techniques revealed large (> 1 μm) linear and branched fibrillar structures with extremely high aspect ratios (Figure 2.5). AFM was used to determine the diameter of a single fiber perpendicular to the image plane (i.e. height), which gave results between 1.5 and 2.2 nm – the predicted width of a rosette. Large aggregated, fibrillar structures created higher order network assemblies that were also observed in some AFM images, consistent with the formation of a gel matrix. TEM micrographs also show the formation of long individual fibers that laterally associate into bifurcating bundles. An electron density profile of one such bundle confirms that the width of an individual fiber is approximately 2 nm (Figure 2.5).

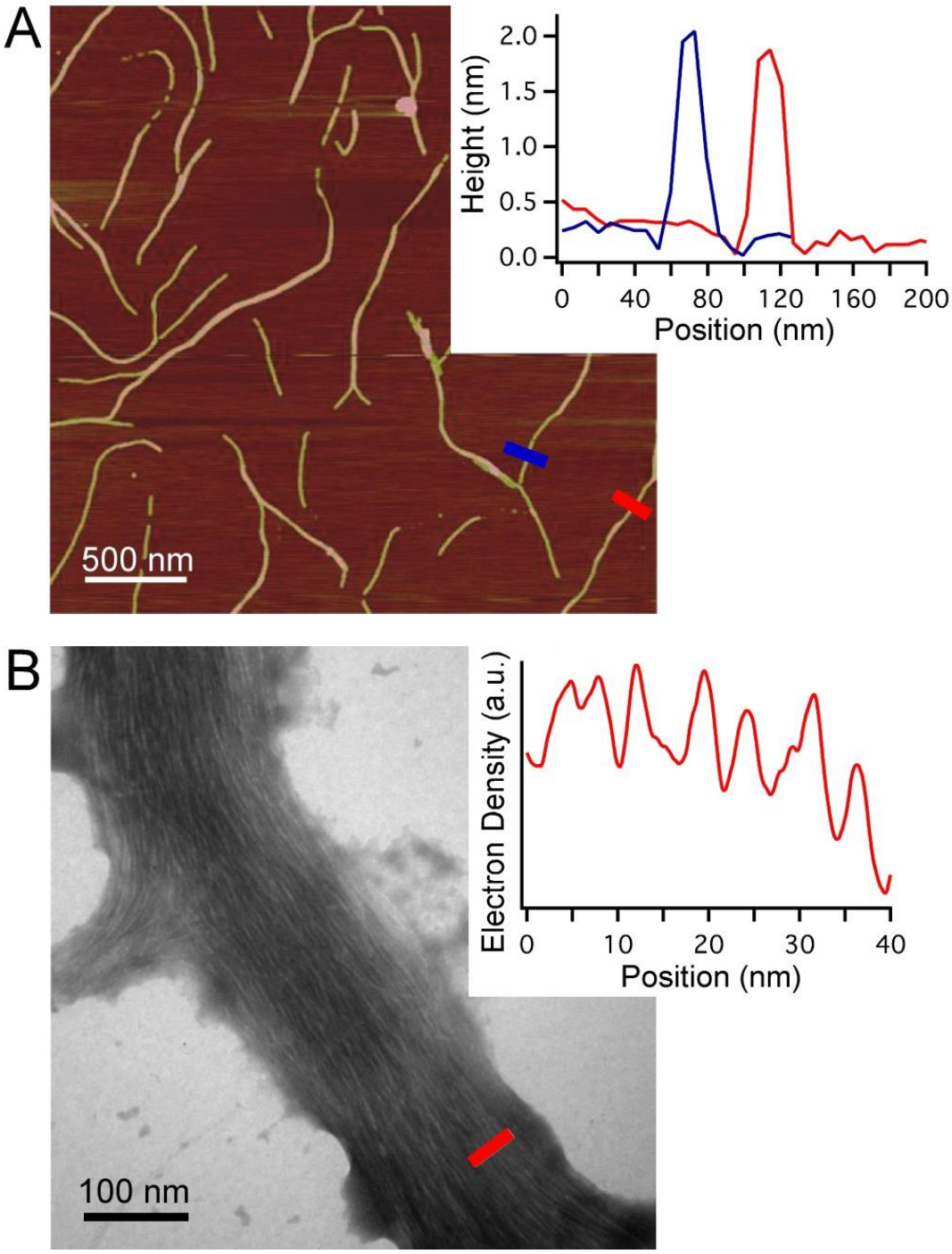


Figure 2.5 AFM and TEM imaging and measurement of TAPAS-Cy supramolecular assemblies. (A) AFM topographic image and (B) TEM micrograph**,** of self-assembled TAPAS-Cy fibers. Insets show the profiles delimited by the red and blue lines in the main panels. The concentration of each monomer used for the assemblies was 5 mM in A, and 10 mM in B.

In a previous study, Fenniri and coworkers demonstrated the formation of micron-length polymer bundles from monomers with similar H-bonding edges [11]. However, the stacking surface of their monomers was larger (i.e. bicyclic) and the formation of polymers longer than 100 nm required covalent tethering of two monomers. Thus, to the best of our knowledge the TAPAS-Cy assemblies are the longest supramolecular polymers generated thus far by untethered, nucleobase-like monomers in water.

While Cy does not have an appreciable excitation coefficient in the UV-vis range (i.e., above 190 nm), TAPAS does with an absorbance maximum at 288 nms at pH 7. The parent heterocycle TAP absorbs maximally at 273 nms under the same conditions and it can be expected that shift in wavelength is a result of conjugation between the ring system and the amide. The UV absorption spectrum of TAPAS changes substantially in the presence of Cy, with the rise of a longer wavelength absorption band (320 nm) at concentrations above 3.5 mM (Figure 2.6), a spectral feature associated with the ordered stacking of chromophors into J-type aggregates [27].

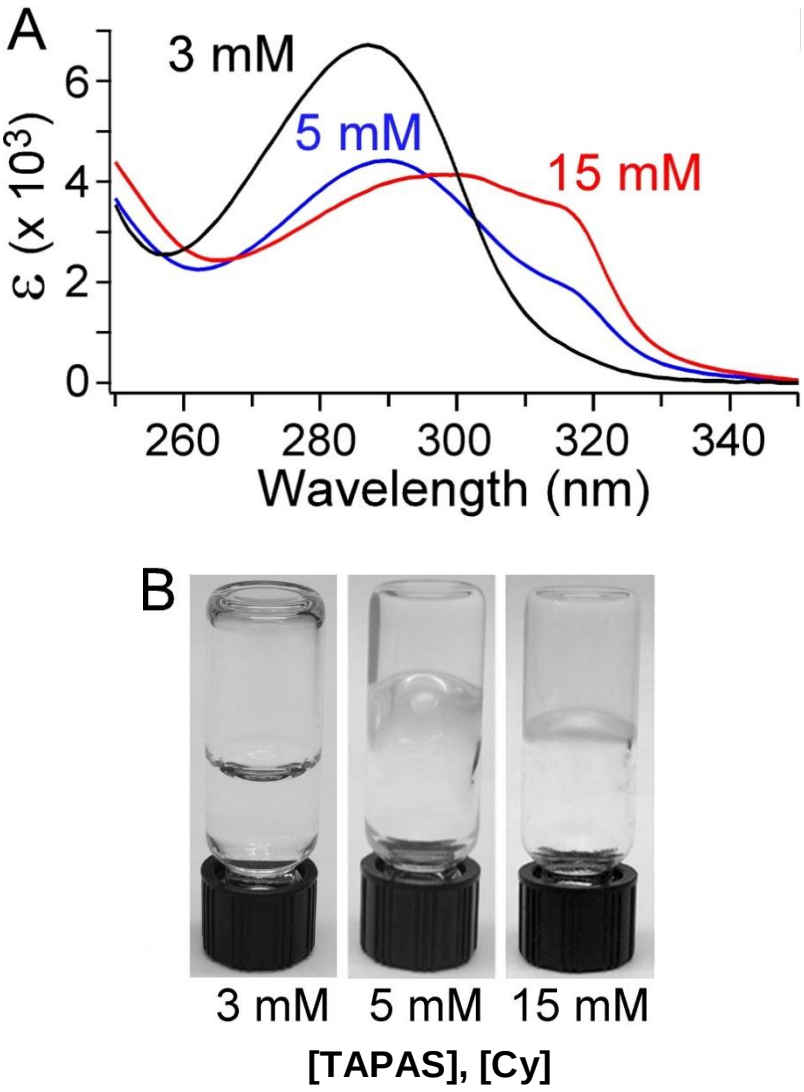


Figure 2.6 UV-vis analyses of TAPAS-CA assemblies at various concentrations. Hydrogel formation with increasing concentration of 1:1 solutions of TAPAS and Cy, at 3, 5 and 15 mM. (B) Inverted-bottle test of gel formation by solutions with the same monomer concentration as in A. Note that the increasing intensity of the band at 320 nm is indicative of ring stacking and correlates with gel rigidity.

2.3.3 TAPAS and Cy Assemble through a Cooperative Assembly Mechanism

NMR was used to follow the association of TAPAS and Cy as a function of concentration to determine their association constant. The ^{1}H spectrum of equimolar mixtures of TAPAS and Cy from 1 to 3 mM exhibited TAPAS chemical shifts identical to a spectrum of TAPAS in the absence of Cy, and with resonance intensities directly proportional to TAPAS concentration (Figure 2.7). At concentrations above 3.5 mM the integrated intensity of TAPAS resonances remained constant at 3.5 mM. The lack of an observed change in TAPAS chemical shifts or signal intensity, or the appearance of new TAPAS resonances, indicates that TAPAS molecules are in slow exchange between their unbound and assembled states (i.e., exchange time constant >100 ms). In these experiments, the TAPAS−Cy assemblies were too large to be observed by solution NMR spectroscopy because of extreme resonance line broadening, consistent with the structures having very large molecular weight, such as those observed by AFM and TEM. The assembly of TAPAS and Cy monomers is consistent with a cooperative assembly mechanism, which is characterized by the coexistence of free monomer and large assemblies (i.e., there are no intermediate structures such as hexad dimers or trimers) [8, 28]. The minimal assembly concentration is thus 3.5 mM at 20 °C. Previously, Adams and coworkers have reported that a low molecular weight gelation system that assembles in organic media through a cooperative assembly mechanism exhibits similar ^{1}H NMR behavior [25]. To the best of our knowledge, such behavior has not been reported for a cooperative supramolecular polymerization system in aqueous solution.

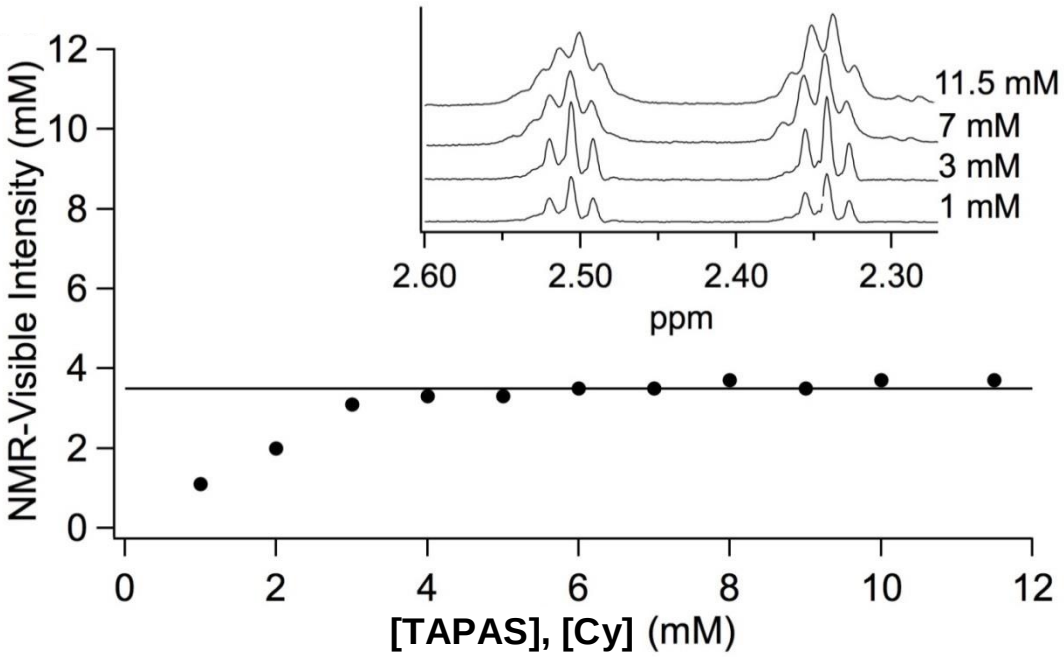


Figure 2.7 Plots of the apparent solution-phase concentration of TAPAS (as determined by methylene ^{1}H NMR resonance integration) vs the actual TAPAS concentration in 1:1 solutions with Cy at pH 7. Horizontal lines indicate the concentration observed for supramolecular assembly at 3.5 mM (i.e., the minimal assembly concentration). The inset show representative spectra of the TAPAS methylene protons. Line broadening in the more concentrated samples likely occurred because the solvent became viscous with gel formation. The peculiar methylene resonance line structure observed at higher concentrations appears to be the result of residual dipolar coupling, which can be considerable in samples containing linear supramolecular assemblies that are partially aligned by the magnetic field.

Additional evidence for a cooperative assembly mechanism was also provided by UV/vis and dynamic light scattering. With increasing concentration of TAPAS and Cy the absorption spectrum of TAPAS does not change until the concentration of monomer is above 3.5 mM. A plot of the relative absorption ratio at 315 nm/288 mn clearly indicates the minimal assembly concentration to be 3.5 mM (Figure 2.8). Additionally, DLS revealed a dramatic increase in scattering intensity when the concentration of each monomer was above 3.5 mM.

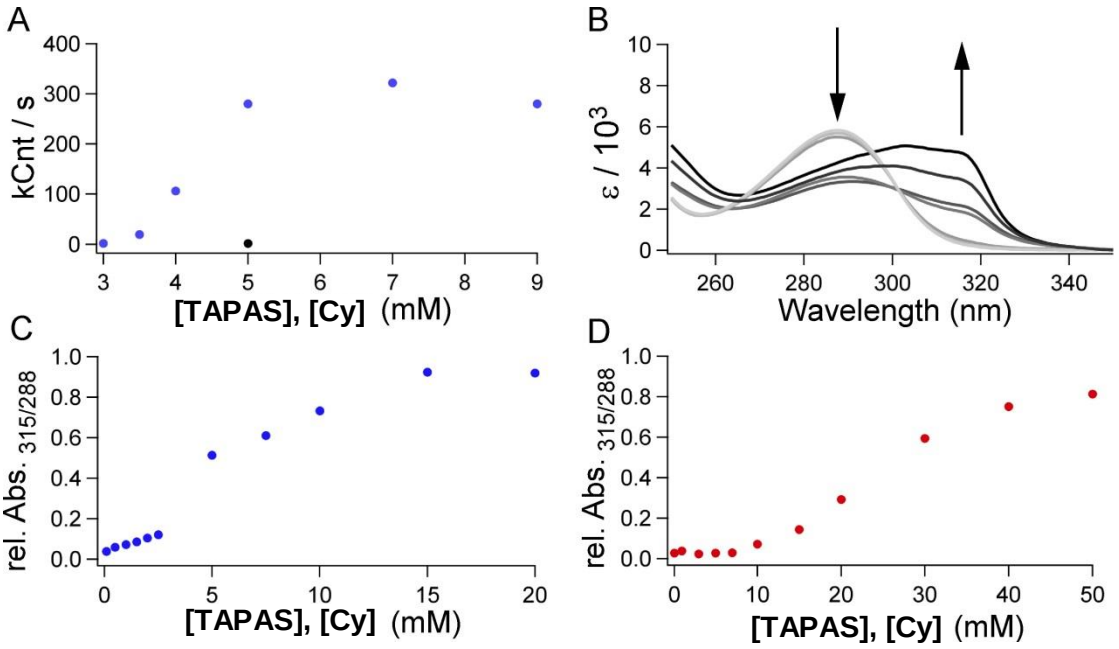


Figure 2.8 DLS and UV absorption analyses of TAPAS-Cy assembly in solutions at various concentrations of 1:1 mixtures of TAPAS:Cy at 20 °C. (A) DLS photon intensity counts indicates a transition to large assemblies above 3 mM. Blue data points represent counts from solutions buffered at pH 7; black data point is from a solution buffered at pH 9. (B) UV absorption spectra of pH 7 solutions from 25 mM to 0.5 mM. Arrows indicate change in spectra as concentration of monomers was increased. (C) Plot of the absorption ratio of $A_{315\ nm}/A_{288\ nm}$ as a function of concentration at pH 7. (D) Same as in c but for solutions buffered at pH 8.

Both variable temperature (VT) ^{1}H NMR and UV/vis demonstrate that the minimal assembly concentration decreases with decreasing temperature (Figure 2.9) indicating an enthalpically driven assembly of monomers. It might be expected that fiber length would have increased with increasing temperature since rosette stacking is primarily driven by the hydrophobic effect. Indeed, Fenniri and coworkers have demonstrated a rosette-based supramolecular polymer that will grow with increasing temperature [29]. However, unlike the TAPAS-Cy system, these reported assemblies are highly stable and show little change in the amount of monomer assembled between 20 and 95 °C.

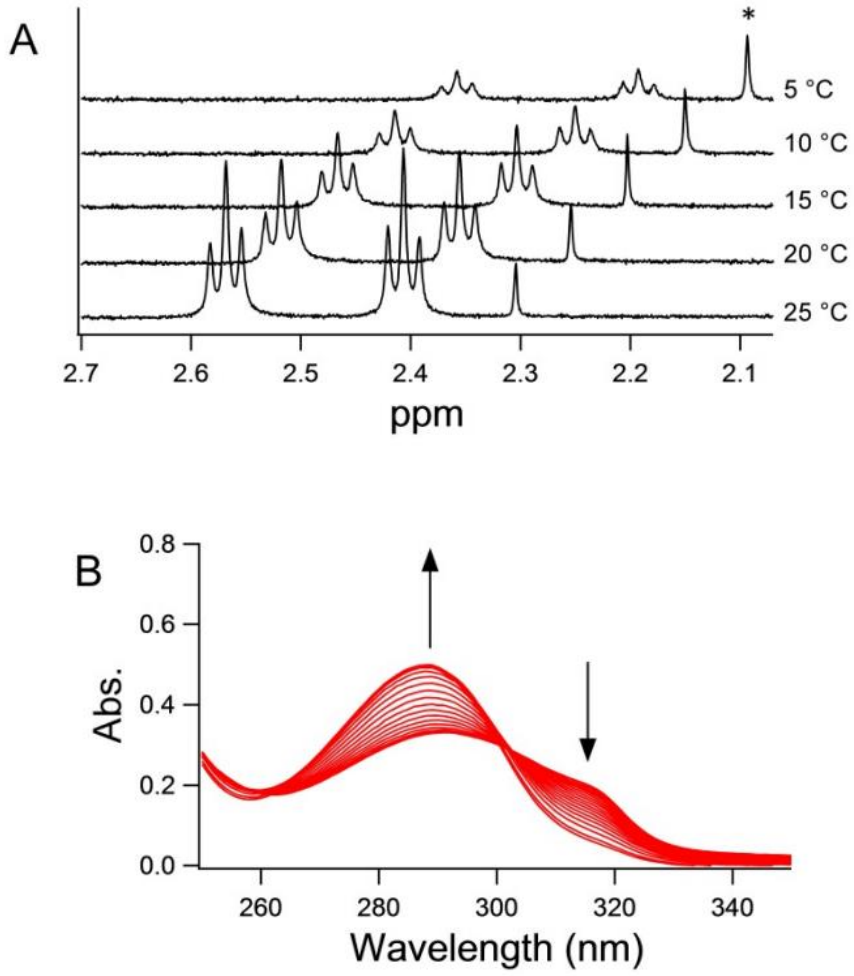


Figure 2.9 Effect of temperature on TAPAS-Cy assembly. (A) NMR spectra of a solution containing a 1:1 mixture of TAPAS:Cy at 5 mM each and various temperatures. Acetate internal standard is labeled (*). (B) UV absorption spectra of a heated 1:1 mixture of monomers at 10 mM each. Arrows indicate the change in intensity upon heating the solution from 15 °C to 50 °C.

The observed cooperative assembly mechanism and thus the coexistence of monomers and very long fibers can be understood as a result of the hydrophobic effect. Theoretical studies estimate the free energy for solvation of a hydrophobic molecule or assembly to be ca. 8 kcal·nm^{-2}·mol^{-1} (calculated for a sphere of radius 0.8 nm) [16]. On the basis of this value, we estimate the free energy cost of exposing the two hydrophobic faces of a single rosette (each face with surface area of 1.7 nm^2) to be on the order of 27 kcal·mol^{-1}. The same positive free energy would apply to the two solvent-exposed ends of a rosette stack and thus may be the origin of the highly cooperative nature of TAPAS-Cy self-assembly. That is, the free energy associated with doubling the average stack length (i.e., combining pairs of existing stacks to eliminate half of the total number of exposed rosette faces) is on the order of -27 kcal·mol^{-1}. In contrast, the free energy of mixing associated with decreasing the concentration of the stacks by a factor of 2 is only +0.3 kcal·mol^{-1}. In view of the large net free energy for stack consolidation, the hydrophobic effect is expected to drive the assembly of rosettes into polymers with lengths that would be limited only by cyclization and the steady-state breakage rates that result from kinetic fluctuations, and thus a considerable factor contributing to the observed cooperative assembly mechanism.

2.3.4 TAPAS-Cy Assemblies are Sensitive to Solution pH and Ionic Strength

At pH 7, free Cy (pK_a 6.9 [30]) exists approximately in equal proportions in its respective neutral and ionized states while TAPAS (pK_a 6.5) has a greater percentage of its heterocycle in the neutral state (ca. 75%). The assembly of a supramolecular structure

formed by TAPAS and Cy would be expected to maintain a net zero charge within the stacked rosettes, otherwise electrostatic repulsion in the low dielectric interior of this structure would prevent formation. In a solution with pH above the pK_a of Cy (i.e., >6.9), three protons per rosette must be abstracted from the solvent upon formation of the supramolecular assembly (an energetically *unfavorable* process). Recently, Bong and coworkers reported that the assembly of Cy with melamine, as either a trivalent MA or Cy derivatives or as the parent heterocycles, can lead to a substantial pK_a shift of the Cy heterocycle [6]. As a test of this prediction, the concentration dependence of the TAPAS-Cy assembly was investigated by ^{1}H NMR spectroscopy at pH 8 where the TAP heterocycle of TAPAS is predominantly neutral and Cy is predominantly charged (Figure 2.10). In this case, the plateau in the integrated resonance intensity was observed at ca. 10 mM TAPAS and Cy, indicating a less favorable assembly than the one observed at pH 7.

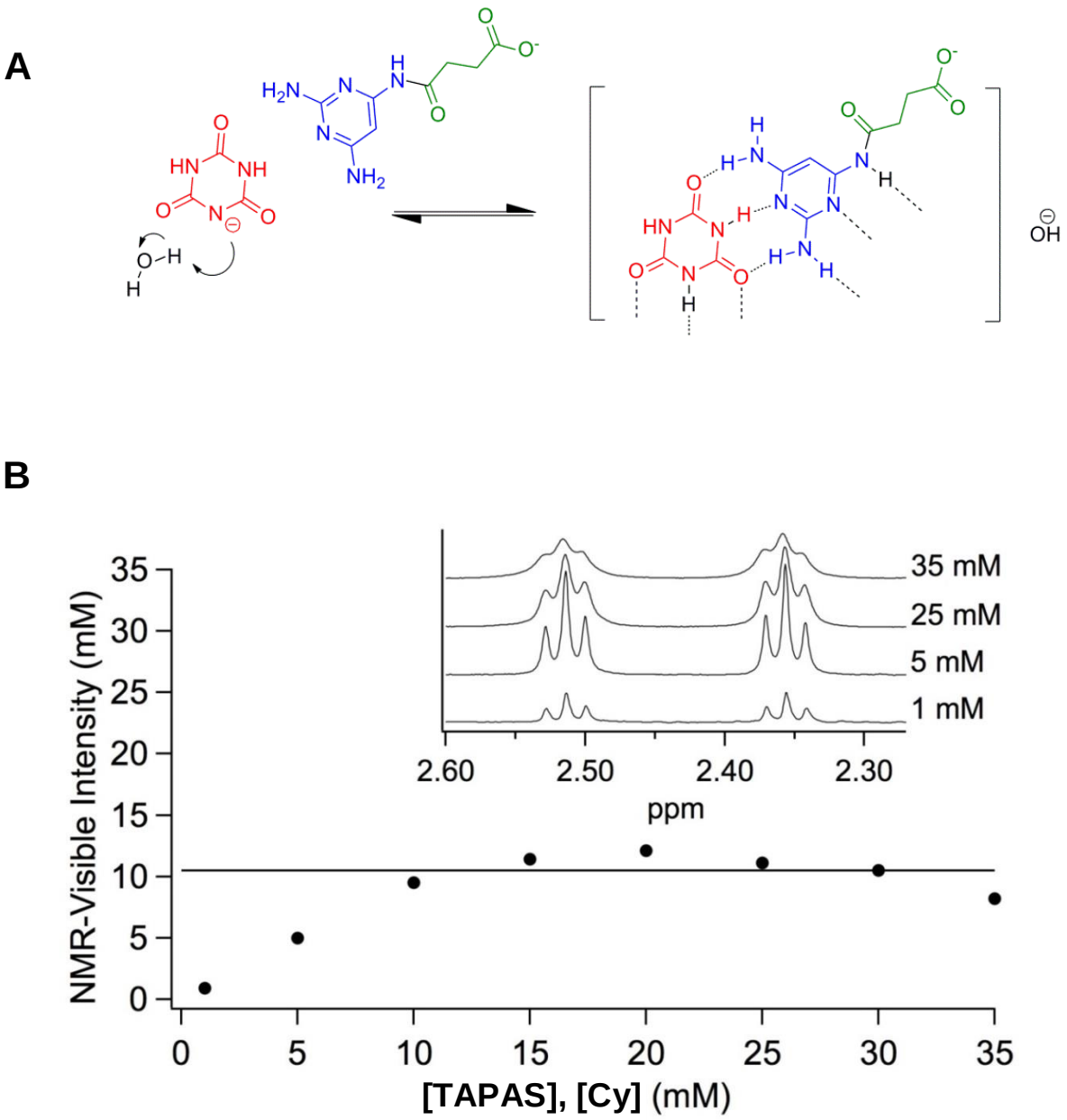


Figure 2.10 Assembly of TAPAS and Cy at pH 8. (A) At pH 8 TAPAS is predominantly neutral (3% charged) and Cy is anionic (9%). In order for TAPAS and Cy to assemble Cy must acquire a proton from solution. (B) Plots of the apparent solution-phase concentration of TAPAS (as determined by methylene ^{1}H NMR resonance integration) vs the actual TAPAS concentration in 1:1 solutions with Cy at pH 8. Horizontal line indicates the 10 mM MAC observed for both monomers under the conditions tested here. The inset show representative spectra of the TAPAS methylene protons. Note that the MAC has increased from where it was at pH 7 (3.5 mM).

The ionic strength of the solution was also found to affect the amount of TAPAS and Cy assembled; demonstrating a third method to control fraction of monomer assembled through environmental changes in addition to temperature and pH. A redshift in the absorbance spectrum of TAPAS is observed as the concentration of NaCl is increased within solutions containing TAPAS-Cy assemblies (Figure 2.11). As discussed above, this spectral change indicates that more TAPAS and Cy are assembling and that the MAC decreases with increasing ionic strength. These findings can be understood in light of reports by the groups of Besenius and Meijer [31, 32], who have shown that charged monomers assemble into supramolecular polymers more favorably in solutions with higher salt due to charge screening along the assemblies periphery; an effect similar to that observed during duplex formation between anionic oligonucleotides [33].

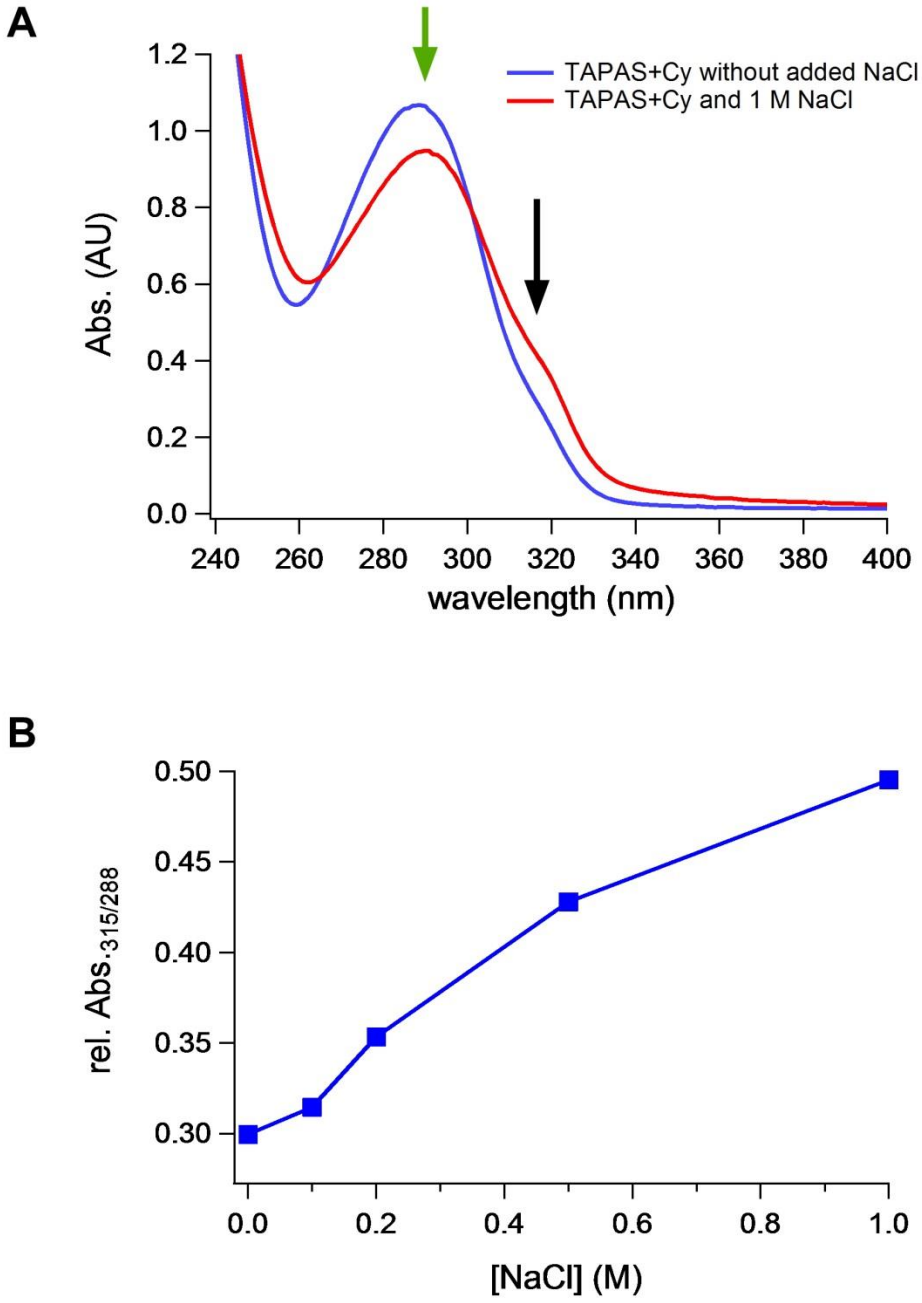


Figure 2.11 Ionic strength dependence of the assembly of TAPAS and Cy. (A) UV absorption spectra of solutions containing TAPAS and Cy with or without the addition of 1 M NaCl. The green arrow indicates the peak at 288 nm of mostly unassembled TAPAS that decreases with increasing NaCl concentration. The black arrow indicates the 315 nm band that is a result of stacking interactions becoming more predominant with increasing NaCl concentration. (B) Plot of the absorption ratio of $A_{315\ nm}/A_{288\ nm}$ as a function of NaCl concentration. All solutions contained 30 mM of both TAPAS and CA at pH 8.

2.3.5 Energetics of TAPAS-Cy Assembly

The free energy associated with TAPAS and Cy monomer incorporation into supramolecular assemblies can be estimated from the free energy of dilution corresponding to the free TAPAS and Cy concentrations that coexist with these assemblies. At 20 °C and pH 7, the 3.5 mM free monomer concentration corresponds to a standard-state free energy of dilution of -3.3 kcal·mol^{-1}. This value is in excellent agreement with free energy values previously determined for interactions between nucleobases within folded RNA structures. Specifically, each H-bond contributes ca. -0.5 kcal·mol^{-1} to the stability of an RNA secondary structure [21], and the stacking of two six-membered rings (derived from the overlap of purine bases on opposite strands of a duplex) contributes -1.8 kcal·mol^{-1} to duplex stability [20, 34]. On the basis of these values, the free energy for the three H-bonds and π–π stacking of each monomer in stacked TAPAS−Cy rosettes is predicted to be−3.3 kcal·mol^{-1}. Moreover, the increase in the MAC at pH 8 can be understood by the addition of if an energetically unfavorable term that can be derived by comparing the abstraction of a proton from water at pH 7 versus pH 8. We estimate the free energy of monomer incorporation into the polymeric assemblies at pH 8 to be -2.7 kcal·mol^{-1} on the basis of the free TAPAS and Cy concentrations of 10 mM. As three protons per rosette must be captured from solution at a free energy that is 1.34 kcal·mol^{-1} greater at pH 8 than at pH 7, this average change in free energy of 0.67 kcal·mol^{-1} per monomer is in excellent agreement with the apparent change in the TAPAS/Cy association free energy from -3.3 kcal·mol^{-1} at pH 7 to ca. -2.7 kcal·mol^{-1} at pH 8 accounting for the 3 fold increase in concentration required for TAPAS-Cy assembly at pH 8.

2.4 Conclusions

The results presented in this chapter illustrate how relatively simple small molecules can efficiently form supramolecular polymers in water if they assemble into structures with hydrophobic surfaces having areas greater than ca. 1 nm^2. The H-bonding and stacking interactions of TAPAS and Cy monomers within our system is similar to the association of nucleobases within RNA and DNA duplexes. However, these associations clearly differ in that the canonical nucleobases (i.e., G, A, C, U, and T) and their corresponding free mononucleosides do not form H-bonded Watson−Crick base pairs in water [35]. A notable exception are G-tetrads formed by guanosine and its derivatives with the aid of cation coordination, which do form polymeric structures in water and can result in gelation [13]. Our explanation for TAPAS-Cy assembly also applies to G-tetrad formation, as the area of a G-tetrad is on the order of 1 nm^2. However, in the case of G-tetrads, two stacked tetrads are observed as an intermediate assembly at pH above 6, and polymer formation is less cooperative than in the TAPAS-Cy system. This tolerance of intermediate assemblies could be due to the smaller stacking area of the G-tetrad and electrostatic repulsions due to coordinated cations in the center of the G-tetrads and phosphate groups on each monomer, as incremental changes in monomer charge can greatly affect the degree of self-assembly in water [28].

The TAPAS-Cy assembly system presented in this chapter demonstrates that the hexameric rosette can be used as a functional architecture to generate hydrogels, which may be useful for soft-material design and applications because of their chemical simplicity. For example, rosette nanotubes have recently been used in the formulation of hydrogel materials that enhance tissue growth [36, 37].

Finally, the inability of the RNA and DNA nucleobases to base-pair and assemble further in water has led to the proposal that the canonical nucleobases are products of evolution [38-40]. The TAPAS-Cy system described here forms polymer assemblies greater than 1 μm in length even at concentrations below 10 mM. Such assemblies contain over 18 000 highly organized monomers. This observation, along with the structural similarity of TAPAS-Cy pairing with Watson−Crick base-pairing, suggests a potential prebiotic mechanism for proto-nucleobase selection from a complex mixture and organization into gene-length polymers even before linkage by a common backbone.

2.5 References

1. Whitesides, G.M. and B. Grzybowski, *Self-assembly at all scales.* Science, 2002. **295**: p. 2418-21.
2. Banerjee, S. and C. Schmuck, *Formation of Hydrogen-Bonded Self-assembled Structures in Polar Solvents.* 2015. **87**: p. 187-225.
3. Prins, L.J., D.N. Reinhoudt, and P. Timmerman, *Noncovalent synthesis using hydrogen bonding.* Angew Chem Int Ed, 2001. **40**: p. 2382-2426.
4. Rehm, T.H. and C. Schmuck, *Ion-pair induced self-assembly in aqueous solvents.* Chem Soc Rev, 2010. **39**: p. 3597-611.
5. Hirschberg, J.H.K.K., et al., *Helical self-assembled polymers from cooperative stacking of hydrogen-bonded pairs.* Nature, 2000. **407**(: p. 167-170.
6. Ma, M.M. and D. Bong, *Determinants of Cyanuric Acid and Melamine Assembly in Water.* Langmuir, 2011. **27**: p. 8841-8853.
7. Aida, T., E.W. Meijer, and S.I. Stupp, *Functional supramolecular polymers.* Science, 2012. **335**: p. 813-7.
8. De Greef, T.F., et al., *Supramolecular polymerization.* Chem Rev, 2009. **109**: p. 5687-5754.
9. Ruckenstein, E. and R. Nagarajan, *Critical micelle concentration. Transition point for micellar size distribution.* J Phys Chem, 1975. **79**: p. 2622-2626.

10. Fenniri, H., et al., *Helical rosette nanotubes: Design, self-assembly, and characterization.* J Am Chem Soc, 2001. **123**: p. 3854-3855.

11. Moralez, J.G., et al., *Helical rosette nanotubes with tunable stability and hierarchy.* J Am Chem Soc, 2005. **127**: p. 8307-8309.

12. Bozsonyi, G., et al., *Water-soluble J-type rosette nanotubes with giant molar ellipticity.* J Am Chem Soc, 2010. **132**: p. 15135-15139.

13. Davis, J.T., *G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry.* Angew Chem Int Ed, 2004. **43**: p. 668-698.

14. Raschke, T.M., J. Tsai, and M. Levitt, *Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solutes in water.* Proc Natl Acad Sci U S A, 2001. **98**: p. 5965-5969.

15. Huang, D.M., P.L. Geissler, and D. Chandler, *Scaling of hydrophobic solvation free energies.* J Phys Chem B, 2001. **105**: p. 6704-6709.

16. Chandler, D., *Interfaces and the driving force of hydrophobic assembly.* Nature, 2005. **437**: p. 640-647.

17. Whitesides, G.M., et al., *Noncovalent synthesis using physical organic chemisistry to make aggregates.* Acc Chem Res, 1995. **28**: p. 37-44.

18. Lehn, J.M., et al., *Molecular ribbons from molecular recognition directed self-assembly of self-complementary molecular-components.* Journal of the Chemical Society-Perkin Transactions 2, 1992: p. 461-467.

19. Bohanon, T.M., et al., *Barbituric-acid 2,4,6-triaminopyrimidine aggregates in water and their competitive interaction with a monolayer of barbituric-acid lipids at the gas-water interface.* Angew Chem Int Ed, 1995. **34**: p. 58-60.

20. Frieir, S.M., et al., *Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCY p, and XCCGGYp helixes.* Biochemistry, 1985. **24**: p. 4533-4539.

21. Klostermeier, D. and D.P. Millar, *Energetics of Hydrogen Bond Networks in RNA: Hydrogen Bonds Surrounding G 1 and U42 Are the Major Determinants for the tertiary Structure Stability of the Hairpin Ribozyme.* Biochemistry, 2002. **41**: p. 14095-14102.

22. Whitesides, G.M., et al., *Noncovalent Synthesis: Using Physical-Organic Chemistry To Make Aggregates.* Acc Chem Res, 1995. **28**: p. 37-44.

23. Hartgerink, J.D., E. Beniash, and S.I. Stupp, *Self-assembly and mineralization of peptide-amphiphile nanofibers.* Science, 2001. **294**: p. 1684-8.

24. Gellert, M., M.N. Lipsett, and D.R. Davies, *Helix Formation by Guanylic Acid.* Proc Natl Acad Sci U S A, 1962. **48**: p. 2013-2018.

25. Hirst, A.R., et al., *Low-molecular-weight gelators: elucidating the principles of gelation based on gelator solubility and a cooperative self-assembly model.* J Am Chem Soc, 2008. **130**: p. 9113-9121.

26. Li, Z., et al., *Structure and gelation mechanism of tunable guanosine-based supramolecular hydrogels.* Langmuir, 2010. **26**: p. 10093-10101.

27. Jelley, E.E., *Spectral absorbance and fluorescence of dyes in their molecular state.* Nature, 1936. **138**: p. 1009-1010.

28. Schaefer, C., et al., *Controlling the Cooperativity in the Supramolecular Polymerization of Ionic Discotic Amphiphiles via Electrostatic Screening.* ACS Macro Letters, 2012. **1**: p. 830-833.

29. Fenniri, H., et al., *Entropically driven self-assembly of multichannel rosette nanotubes.* Proc Natl Acad Sci U S A, 2002. **99**: p. 6487-6492.

30. Jang, Y.H., et al., *Acid Dissociation Constants of Melamine Derivatives from Density Functional Theory Calculations.* J Phys Chem B, 2009. **113**: p. 13036-13040.

31. Besenius, P., et al., *Controlling the growth and shape of chiral supramolecular polymers in water.* Proc Natl Acad Sci U S A, 2010. **107**: p. 17888-17893.

32. Schaefer, C., et al., *Controlling the Cooperativity in the Supramolecular Polymerization of Ionic Discotic Amphiphiles via Electrostatic Screening.* ACS Macro Letters, 2012. **1**: p. 830-833.

33. SantaLucia, J. and D. Hicks, *The thermodynamics of DNA structural motifs.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 415-440.

34. Mohan, S., et al., *RNA Tetraloop Folding Reveals Tension between Backbone Restraints and Molecular Interactions.* J Am Chem Soc, 2010. **132**: p. 12679-12689.

35. Ts'o, P., I. Melvin, and A. Olson, *Interaction and association of bases and nucleosides in aqueous solutions.* J Am Chem Soc, 1963. **85**: p. 1289-1296.

36. Zhang, L., et al., *Arginine-glycine-aspartic acid modified rosette nanotube-hydrogel composites for bone tissue engineering.* Biomaterials, 2009. **30**: p. 1309-1320.

37. Beingessner, R.L., et al., *Bioactive Rosette Nanotubes for Bone Tissue Engineering and Drug Delivery.* 2013: p. 313-357.

38. Hud, N.V. and F.A.L. Anet, *Intercalation-mediated synthesis and replication: a new approach to the origin of life.* J Theor Biol, 2000. **205**: p. 543-562.

39. Hud, N.V., et al., *The origin of RNA and 'My Grandfather's Axe'.* Chem Biol, 2013. **20**: p. 466-474.

40. Cafferty, B.J. and N.V. Hud, *Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology.* Curr Opin Chem Biol, 2014. **22**: p. 146-157.

CHAPTER 3

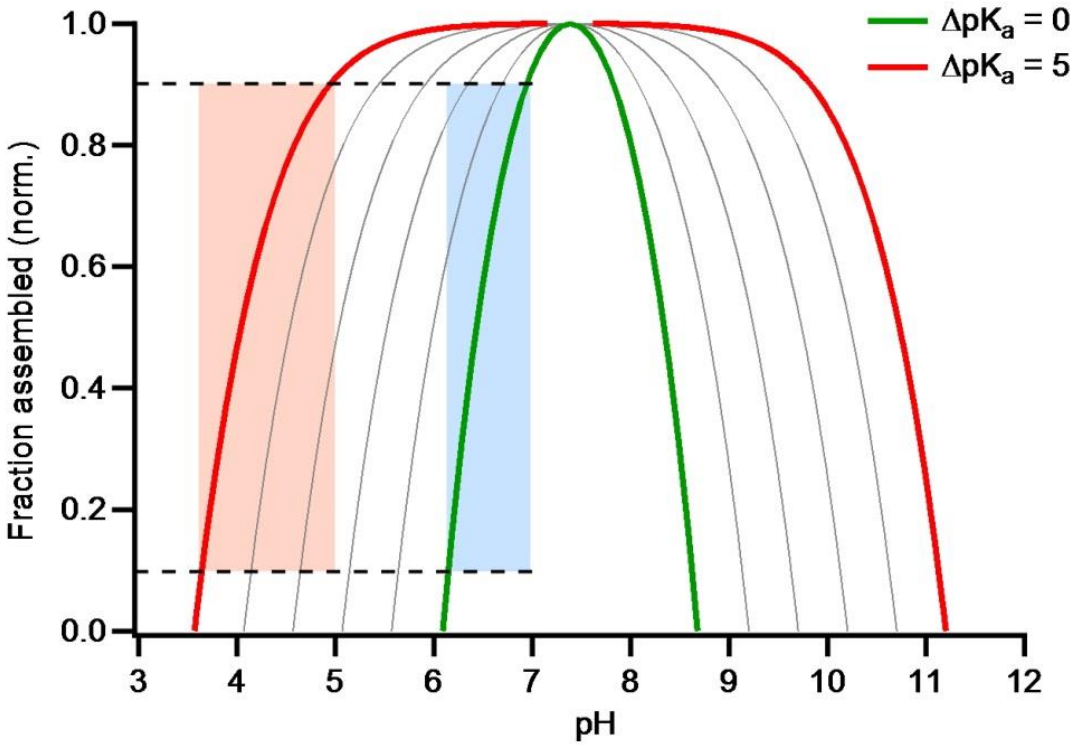
PRECISE pH CONTROL OF SUPRAMOLECULAR POLYMERS AND HYDROGELS: MONOMER pK_a MATCHING AND POLYMER HETEROGENEITY[c]

3.1 Introduction

Precise control over the self-assembly of small molecules in water underlies developmental strategies for many promising materials [1, 2], including those for use in drug delivery and tissue engineering [3, 4]. Control of supramolecular polymer assembly and hydrogel formation by low molecular weight gelators can be achieved by using monomers with intermolecular interactions that are altered by environmental changes, such as pH, temperature and the addition of other molecules or ions [5-10]. Unfortunately, it is the case with many low molecular weight gelators that the same interactions that enable hydrogelation also lead to the formation of precipitates and, overtime, gel collapse [11, 12]. Therefore, new approaches that allow precise control over the supramolecular assembly process and increase gel lifetime are critical for supramolecular systems to be integrated into functional materials.

c. This chapter was adapted from previously published work and is reproduced with permission. Cafferty, BJ; Avirah, RR; Schuster, GB; Hud, NV "Ultrasensitive pH control of supramolecular polymers and hydrogels: pK_a matching of biomimetic monomers." *Chem. Sci.* (2014).

An appealing approach to the design of self-assembling structures for biological applications is the creation of supramolecular polymers that respond in a highly non-linear manner to modest pH changes within the physiological pH range [13]. Bidirection pH control of self-assembly has been demonstrated for monomers that contain acidic and basic groups that promote assembly through charge pairing within the pH range between the pK_as of the charged groups [14-16]. The pH range over which such an assembly is stable and responsive, typically defined as $\Delta pH_{10\text{-}90}$ (where 10-90% of the assemblies are present in solution), is coupled to the pK_as of the acidic and basic recognition elements. The response of a system to changes in pH can therefore be enhanced by minimizing the ΔpK_a of the recognition elements (Figure 3.1). While bidirectional systems commonly employ amines and carboxylic acids to enable charge pairing, these functionalities do not have pK_as near pH 7 and have a ΔpK_a of at least 5, making these systems mostly insensitive to pH changes near neutrality (Figure 3.1). It is therefore reasonable to predict that maximum pH sensitivity for a supramolecular polymerization process that responds to pH changes that are physiologically relevant could be achieved with assemblies that employ recognition elements with pK_as matched near pH 7 (Figure 3.1).



$$\mathbf{NFA} = \frac{\mathbf{A}_{\mathrm{tot}} - \sqrt{K_{\mathrm{sp}}\left(1 + 10^{\mathrm{pH} - \mathrm{p}K_{\mathrm{BH}\oplus}}\right)\left(1 + 10^{\mathrm{p}K_{\mathrm{AH}} - \mathrm{pH}}\right)}}{\mathbf{A}_{\mathrm{tot}} - \sqrt{K_{\mathrm{sp}}}\left(1 + 10^{\mathrm{p}K_{\mathrm{AH}} - \mathrm{p}K_{\mathrm{ave}}}\right)}$$

Figure 3.1 Theoretical limits for pH-dependent bidirectional assembly systems with ΔpK_as ranging from 0 to 5. The sensitivity of the assembly to change in pH, $\Delta pH_{10\text{-}90}$, is highlighted in blue for $\Delta pK_a = 0$ (pK_a matched), and pink for $\Delta pK_a = 5$ (the minimal ΔpK_a for assembly systems that use amines and carboxylic acids for charge pairing). Note that the pH range over which assemblies are present increases with increasing ΔpK_a; the system with $\Delta pK_a = 0$ shows the greatest sensitivity to changes in pH. Curves of fraction assembled as a function of pH were generated by taking the product of the pairing monomers at pH 7.4 (the ΔpK_a midpoint) as the solubility constant (Ksp) of each system. This constant and the calculated monomer ionization equilibrium were then used to calculate the maximum expected assembly as a function of pH (normalized to fraction assembled at pH 7.4, for pK_{ave}) using the equation shown under the plot. Calculations were performed using idealized model systems that all have a minimum assembly concentration (MAC) of 15 mM in each monomer at pH 7.4, and a total sample concentration of 35 mM in each monomer. NFA stands for normalized fraction assembled. See section 3.5 below for full derivation and example.

Work on the coassembly of monomers containing the triazines melamine and cyanuric acid (Cy), or their related pyrimidines triaminopyrimidine (TAP) and barbituric acid, as complementary hydrogen-bonding recognition units has fundamentally advanced our understanding of artificial self-assembly [17, 18], particularly in organic solvents where initial investigations were performed. As discussed in the preceding chapter, the addition of a hydrophilic side chain to one of these recognition elements promotes the formation of water soluble supramolecular assemblies [19, 20]. Of these heterocycles, TAP and Cy are unique as they are complementary and pK_a matched near neutrality. Critically, Cy is predominantly negatively charged above this shared pK_a, while TAP is positively charged below it, precisely the properties necessary for a self-assembling system with ultimate possible sensitivity to pH.

Here, reversible supramolecular polymers formed from a mixed system of monomeric units tagged with TAP and Cy recognition elements are evaluated for their sensitivity to pH change. In the previously discussed assembly system TAP was modified with a charged anionic peripheral group (TAPAS), however this made TAP poorly soluble at lower pH where the molecule is a zwiterion. To develop a water soluble monomer that retained solubility, as well as to ascertain if Cy could function in the role TAP had for TAPAS, monomers composed of Cy substituted with a hexanoic acid (CyCo6) or a butyric acid (CyCo4) were prepared (Figure 3.2) [21]. Mimicking nucleic acid base pairing, hydrogen bonding directs the assembly of TAP and the CyCo monomers into hexameric rosette structures, and hydrophobic interactions cause these rosettes to stack into supramolecular polymers that can then non-covalently crosslink to form a hydrogel matrix. Supramolecular assemblies were found to be most stable when

composed of all three monomers and to exhibit the maximal sensitivity to change in pH for a supramolecular polymerization system.

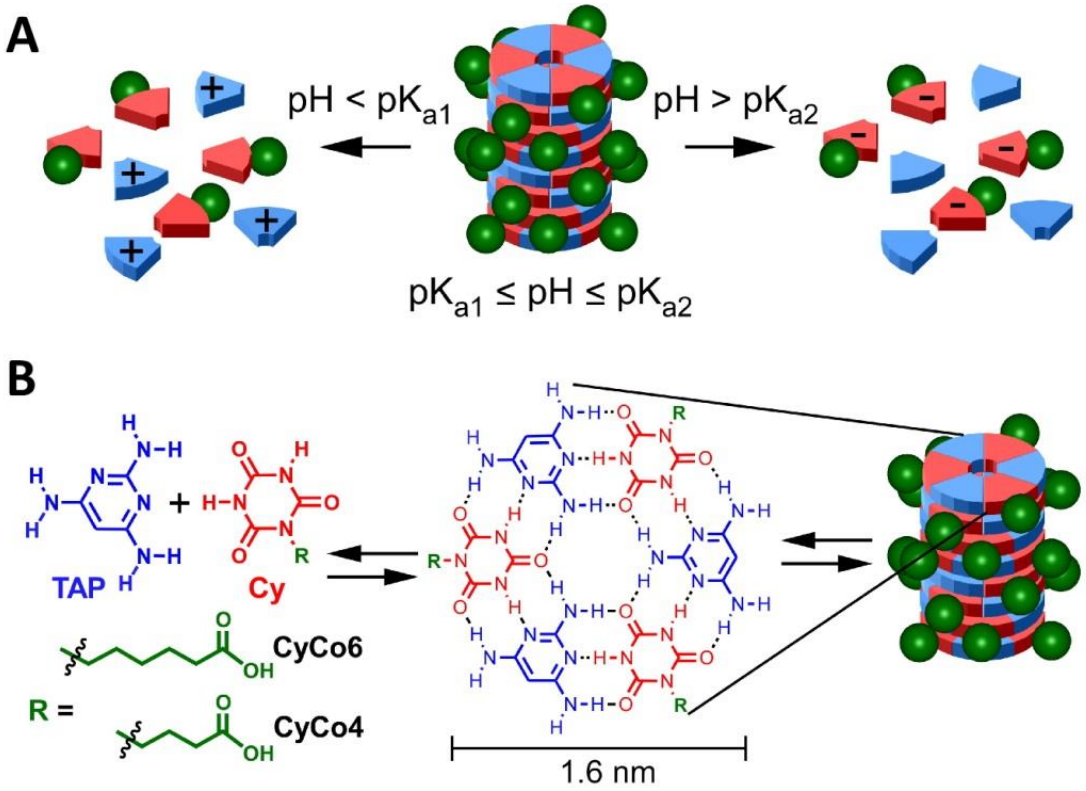


Figure 3.2 Graphical representation of TAP-CyCo assembly system. (A) Graphical representation of pH-mediated assembly and disassembly by monomers with pK_a-matched recognition units. (B) Chemical structures of triaminopyrimidine (TAP), cyanuric acid (Cy, R = H), CyCo6, CyCo4, and proposed rosette and stacked rosette assemblies.

3.2 Experimental Procedures

3.2.1 Materials

2,4,6-triaminopyrimidine (TAP) and cyanuric acid (Cy) were purchased from Acros Organic. All materials were used as received.

3.2.2 Synthesis and Characterization

Synthesis of 1-(5-Carboxypentyl)-1,3,5-triazin-2,4,6-trion (CyCo6) and 1-(3-Carboxypropyl)-1,3,5-triazin-2,4,6-trion (CyCo4) was performed by following and modifying the procedure reported by Hager *et al.* [21].

CyCo6: ^{1}H NMR (500 MHz, DMSO-d_6): δ = 1.27 (m, 2H; CH_2), 1.50 (m, 4H; CH_2), 2.19 (t, J = 7.5 Hz, 2H; CH_2CO), 3.61 (t, J = 7.5 Hz, 2H; CH_2N), 11.63 ppm (brm, NH); ^{13}C NMR (125 MHz, DMSO-d_6): δ = 24.1, 25.6, 27.05, 33.5, 40.2, 148.6, 149.8, 174.4 ppm; HRMS *m/z* calcd. for $[M - H]^-$, 242.0777; found, 242.0785.

CyCo4: ^{1}H NMR (500 MHz, DMSO-d_6): δ = 1.75 (m, 2H; CH_2), 2.24 (t, J = 7.5 Hz, 2H; CH_2CO), 3.68 (t, J = 7 Hz, 2H; CH_2N), 11.63 ppm (brm, NH); ^{13}C NMR (125 MHz, DMSO-d_6): δ = 23.2, 31.4, 40.5, 149.2, 150.44, 174.5 ppm; HRMS *m/z* calcd. for $[M - H]^-$, 214.0464; found, 214.0451.

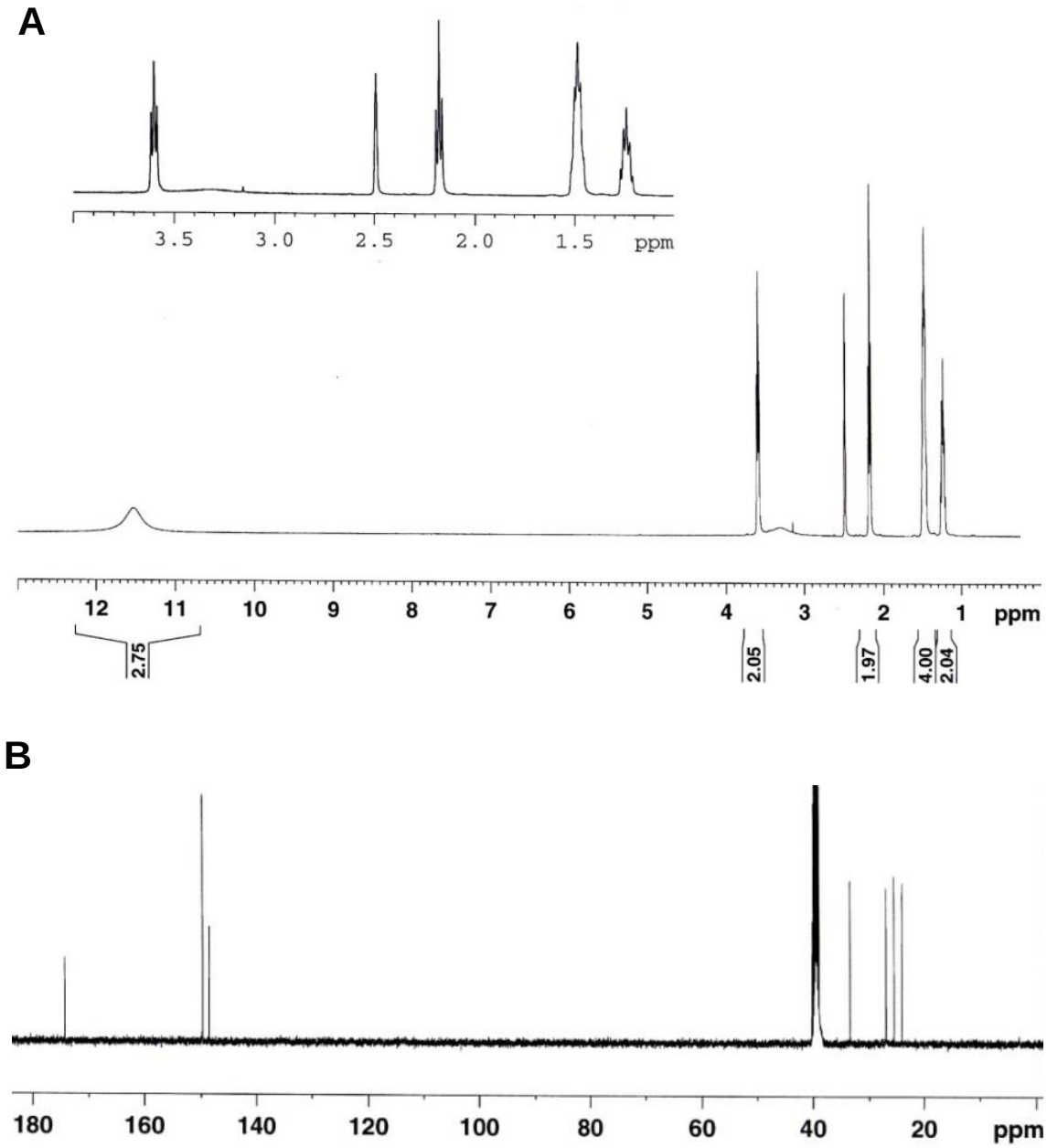


Figure 3.3 NMR characterization of CyCo6**. (**A) ^{1}H NMR spectrum of CyCo6 in DMSO-d_6. (B) ^{13}C NMR spectrum of CyCo6 in DMSO-d_6.

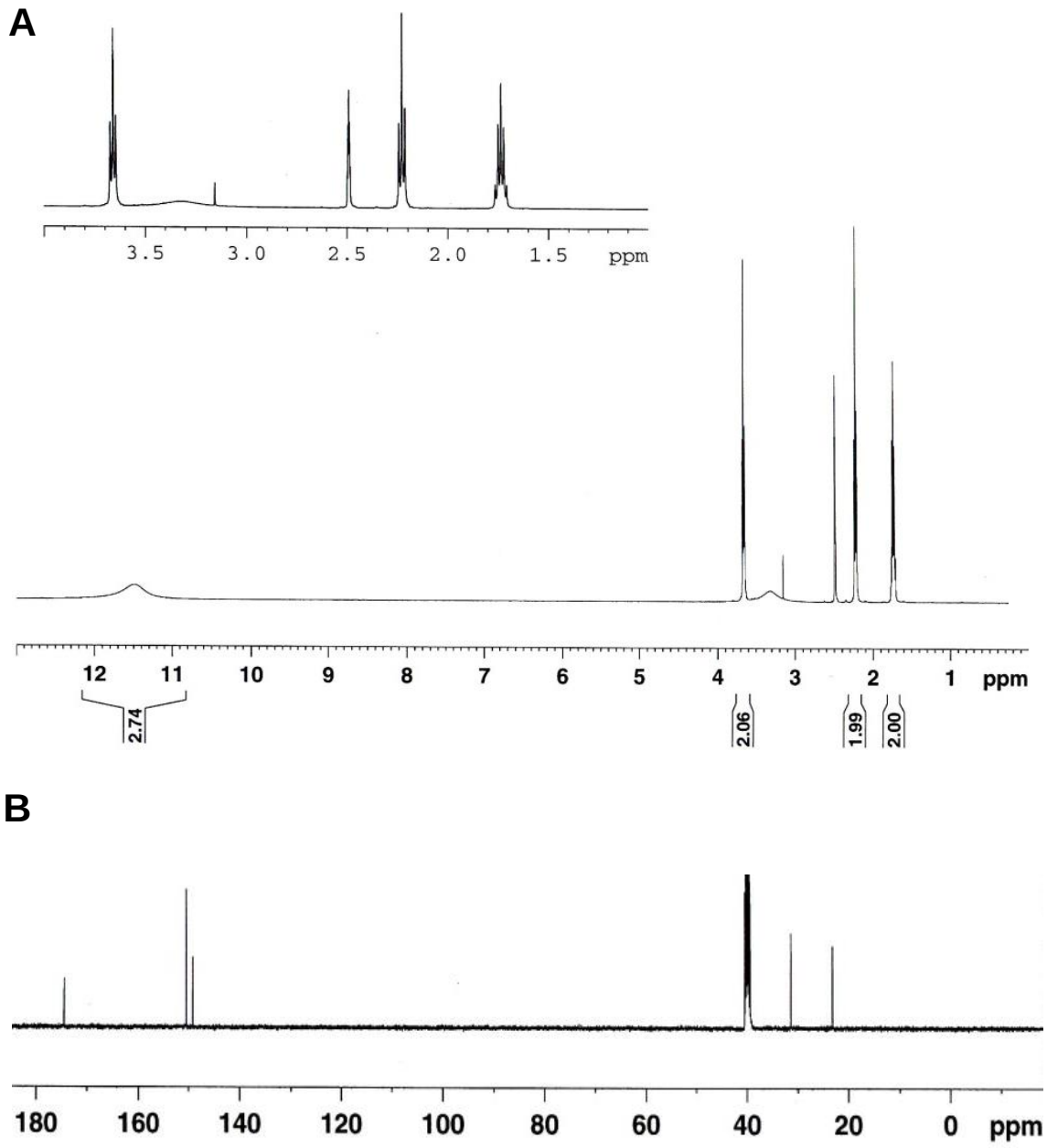


Figure 3.4 NMR characterization of CyCo4**. (**A) ^{1}H NMR spectrum of CyCo4 in DMSO-d_6. (B) ^{13}C NMR spectrum of CyCo4 in DMSO-d_6

3.2.3 Sample Preparation

Assemblies were formed by adding CyCo6 and/or CyCo4 to a solution containing an equivalent amount of TAP and 200 mM sodium phosphate buffered at pH 7; pH was adjusted with NaOH or HCl. Unless otherwise noted, assembly solutions were equilibriated and analyzed at 20 °C. CyCo4 and CyCo6 purity were determined by NMR and LCMS. NMR analysis for molecular characterization (Figure 3.3 and 3.4) and precipitate composition was performed in DMSO-d_6 on a Varian Mercury 400 MHz NMR. HRMS was performed using a Waters Synapt G2. pH measurements were taken with a VWR 8100 pH meter equipped with an InLab semi-micro combination electrode.

3.2.4 Instrumental Methods

Spectroscopic analysis of the assemblies was performed using UV-vis and ^{1}H NMR. UV-vis analysis was carried out on an Agilent 8453 spectrophotometer equipped with an 89090A temperature controller. Cells of different path lengths (0.1 and 0.01 mm) were used depending on the concentration of the sample to maintain an optical density below 1.2. ^{1}H NMR spectra were collected on a Bruker DRX-500 500 MHz NMR and were the sum of 64 transients. Assembly solutions investigated by NMR contained 90% H_2O and 10% D_2O and were observed using the WATERGATE pulse sequence. Cytosine was used as an internal standard at a starting concentration of 25 mM, cytosine did not show any indication of interacting or being incorporated within the assemblies.

AFM imaging was performed on freshly cleaved mica that was pre-activated with $MgCl_2$ with 1-2 h incubation. The mica substrate was rinsed with water and dried under N_2. Solutions containing the assemblies were incubated on ice just prior to deposition. A

2 µl sample of the assembly solution was spin coated for 30 sec at 2500 rpm and dried with N_2 gas. AFM imaging was performed with a Nanoscope IIIa (Digital Instruments) in tapping mode, using Si tips (Vistaprobes, 48 N/m).

Rheological measurements were carried out on a Physica MCR 501 rheometer (Anton Paar). The storage modulus, G′, and loss modulus, G″, were measured in oscillatory tests at a constant angular frequency of 1 rad/s while sweeping the strain. Frequency scans were performed under a strain of 1.0%. All measurements were temperature controlled with a Peltier plate at 20 °C.

3.2.5 TAP and CyCo pK_a Determination

The pK_as of the heterocyclic rings of CyCo4 and CyCo6 were determined by ^1H NMR because the heterocycle does not have appreciable UV-vis absorbnce. Change in the chemical shift of the methylene proton of the aliphatic acid tails of CyCo4 and CyCo6 closest to the heterocycle were plotted and a sigmodal fit was used to determine the pK_a of the Cy moiety. CyCo6, 7.3; CyCo4 7.3 (Figure 3.5).

The pK_a of the TAP was determined by following change in absorbance at 272 nm (which is the maximal absorbance of the protonated pyrimidine ring) in the same buffer conditions used for the assembly studies, 200 mM sodium phosphate buffer at 20 °C. The first derivative of a sigmodal fit revealed that the pK_a of TAP in phosphate buffer is 7.5. pH was varied by addition of 1 M NaOH or HCl and a pH probe was used to verify solution pH. The pK_a was also determined in a Tris-Citrate buffer. In this case the pK_a was found to be closer to pH 7. This indicates that the phosphate buffer may be interacting with TAP (as it is known to do with melamine) causing a pK_a shift.

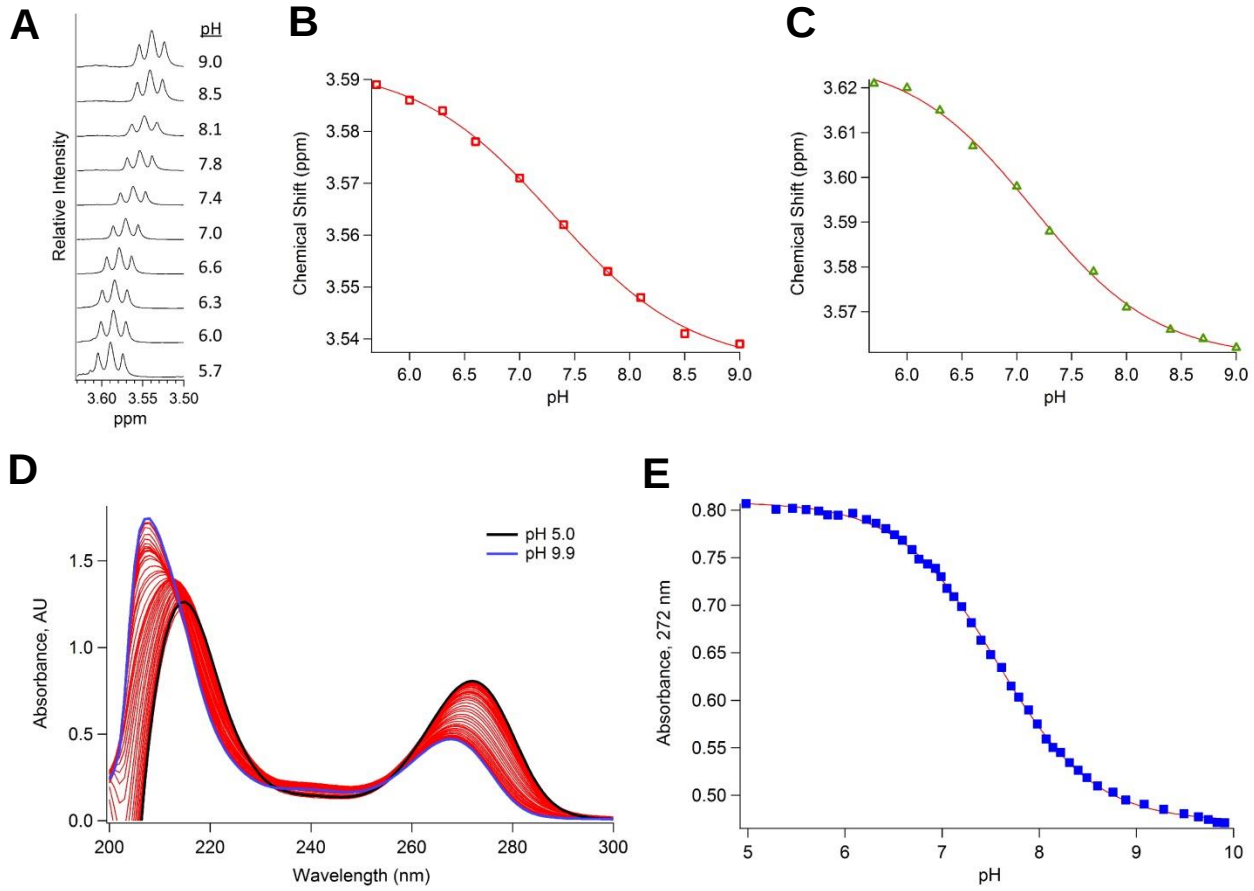


Figure 3.5 pK_a determination of the Cy heterocycle for CyCo6 and CyCo4 by ^1H NMR and for TAP by UV-vis in 200 mM sodium phosphate buffer at 20 °C. (A) Representative ^1H NMR spectra of the methylene proton peak used to determine the pK_a of CyCo6. (B) Plot of CyCo6 methylene peak position as a function of pH. (C) Plot of CyCo4 methylene peak position as a function of pH. (D) UV-vis spectra of TAP (50 uM) at various pHs between 5.0 and 9.9. (E) Change in absorbance of TAP at 272 nm (where the protonated form of TAP absorbs maximally) as a function of pH. Sigmoidal fits (shown) reveal pK_as of 7.3 for CyCo6, 7.2 for CyCo4 and 7.5 for TAP.

3.3 Results and Discussion

3.3.1 CyCo-TAP Co-Assemble into Supramolecular Polymers

A shear-thinning, thermally reversible supramolecular hydrogel forms spontaneously when one equivalent of TAP is added to a 20 mM solution of CyCo6 (0.7% by weight in total monomer), at pH 7 and 20 °C. The UV/vis absorption spectrum of TAP in the gel phase is red shifted relative to the spectrum of free TAP, indicative of intermolecular π-π electron interactions that are typical of J-type associations by planar molecules (Figure 3.6). Gelation is not observed below pH 6 or above pH 8 (Figure 3.6); demonstrating the pH sensitivity of the gel-forming process. At pH 7 and pH 8 TAP and CyCo6 remain in solution for times greater than six months, but precipitation occurs at pH 6 within minutes. Under the same conditions CyCo4 forms a precipitate immediately when mixed with TAP (Figure 3.7). Incorporation of a cogelator has been shown to enhance the lifetime of the hydrogel phase [22]. We hypothesized that non-uniform structures formed by the inclusion of closely related CyCo6 as a co-monomer into TAP-CyCo4 assemblies would frustrate regular interactions between linear stacks of monomers. As predicted, mixtures of CyCo4 and CyCo6 form hydrogels with TAP whose gel lifetime at pH 7 increases as the ratio of CyCo6 to CyCo4 increases (Figure 3.8). For example, a solution containing TAP, CyCo4 and CyCo6 at a ratio of 1:0.8:0.2, respectively, forms a gel that begins to precipitate within five minutes while a solution at a 1:0.33:0.67 ratio in each molecule, respectively, forms a gel that is stable for greater than 5 hours. At pH 6 the trimolecular assemblies are

stable against precipitation for over an hour, while the bimolecular assemblies precipitate rapidly, demonstrating a synergistic relationship among the monomers.

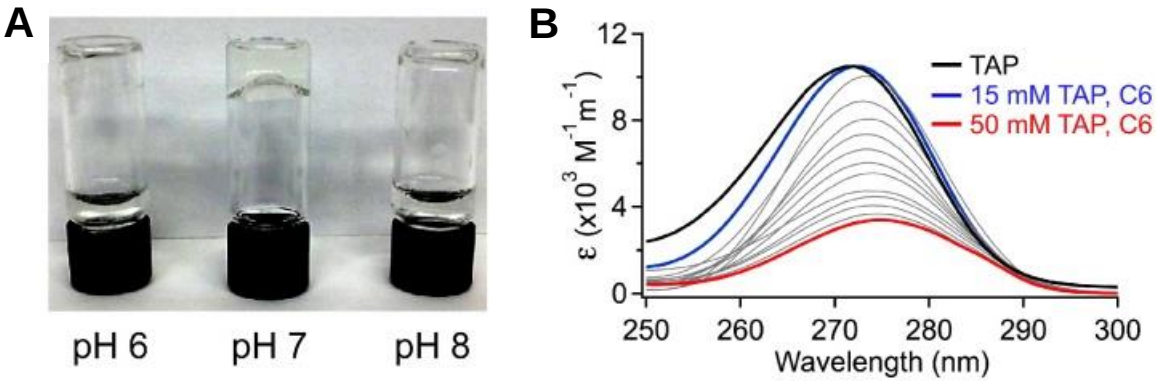


Figure 3.6 Inverted bottle test and UV-vis analysis of TAP-CyCo6 assemblies. (A) Inverted bottle test showing gelation of TAP with CyCo6 at pH 7 and no gelation at pH 6 or pH 8. All three solutions are 20 mM in both monomers. (B) UV spectra of solutions of TAP alone and TAP with CyCo6 at equimolar concentrations from 15 to 50 mM.

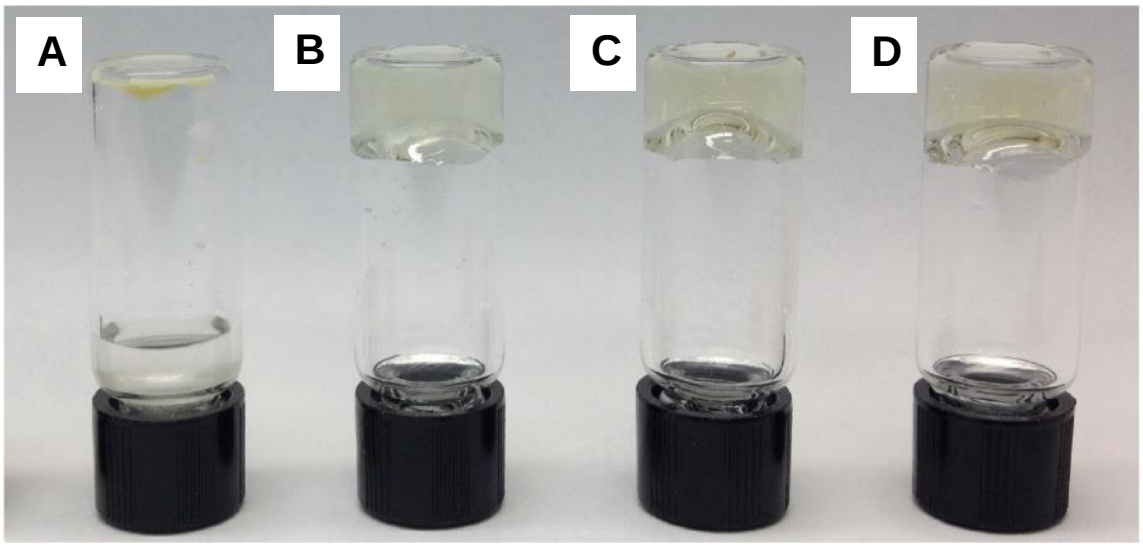


Figure 3.7 Gel-inversion tests of solutions containing TAP, CyCo6 and CyCo4 at various ratios of CyCo6 and CyCo4. Each vial contains 20 mM TAP and 20 mM modified-Cy species buffered at pH 7 at 20 °C. Solutions contained (A) CyCo4 and TAP (precipitate, no gelation), (B) same as A, but with an additional 1 M NaCl which shows evidence of precipitation within 5 minutes, (C) TAP, CyCo4 and CyCo6 in a 1:0.5:0.5 ratio, respectively (no additional NaCl) which begins to precipitate after 30 minutes, and (D) TAP, CyCo4 and CyCo6 in a 1:0.8:0.2 ratio, respectively, which begins to precipitate after 5 minutes. Note that the rate of precipitation is decreased at lower temperature, for example precipitation began after 2 hours for gels described in C when stored at 4 °C.

The nature of TAP-CyCo assemblies at the molecular level was investigated using spectroscopic and microscopic techniques. The UV-vis spectrum of samples containing 1:1 TAP-CyCo6 and TAP-CyCo4-CyCo6, at various ratios where the total Cy containing monomer is equal to the concentration of added TAP, revert to that of free TAP upon dilution from 40 mM to 15 mM (in TAP) (Figure 3.8), indicating the absence of intermolecular stacking interactions at concentrations below 15 mM. AFM imaging of the TAP-CyCo6 and the trimolecular gel phase shows micron-length linear structures with heights of approximately 2 nm (Figure 3.9). These structures are consistent with the expected width of the stacked rosette assembly. As can be seen from Figure 3.9B, the supramolecular polymers are so long that most of the ends cannot be identified indicating that the average length is greater than 5 μm. The observation of extended networks of fibers by AFM is also consistent with the gelation properties of the assemblies.

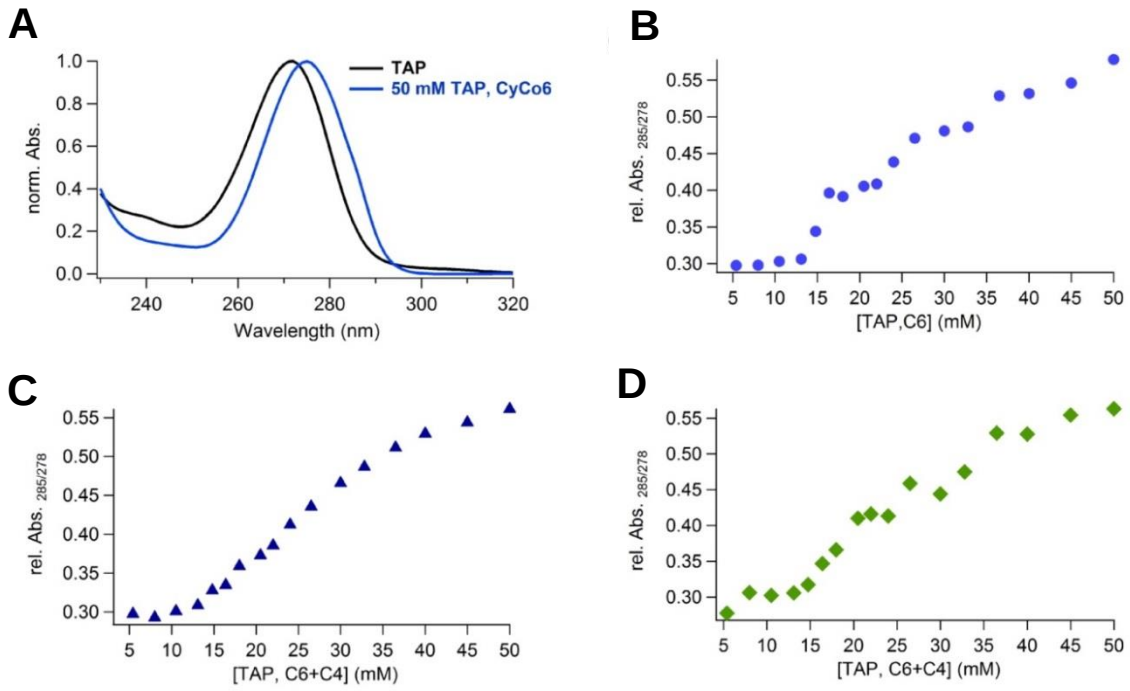


Figure 3.8 UV absorption analysis of TAP, CyCo4 and CyCo6 at various concentrations. (A) Spectra of TAP alone (black) and TAP in a solution with CyCo6 at 50 mM in each monomer (blue) where the absorption maxima have been normalized to unity. (B-C) Plot of the absorption ratio of A285 nm/A278 nm as a function of monomer concentration in (B) 1:1 mixtures of TAP:CyCo6, (C) 1:0.5:0.5 mixtures of TAP:CyCo4:CyCo6, and (D) 1:0.33:0.66 mixtures of TAP:CyCo4:CyCo6. All solutions were buffered at pH 7 and analyzed at 20 °C.

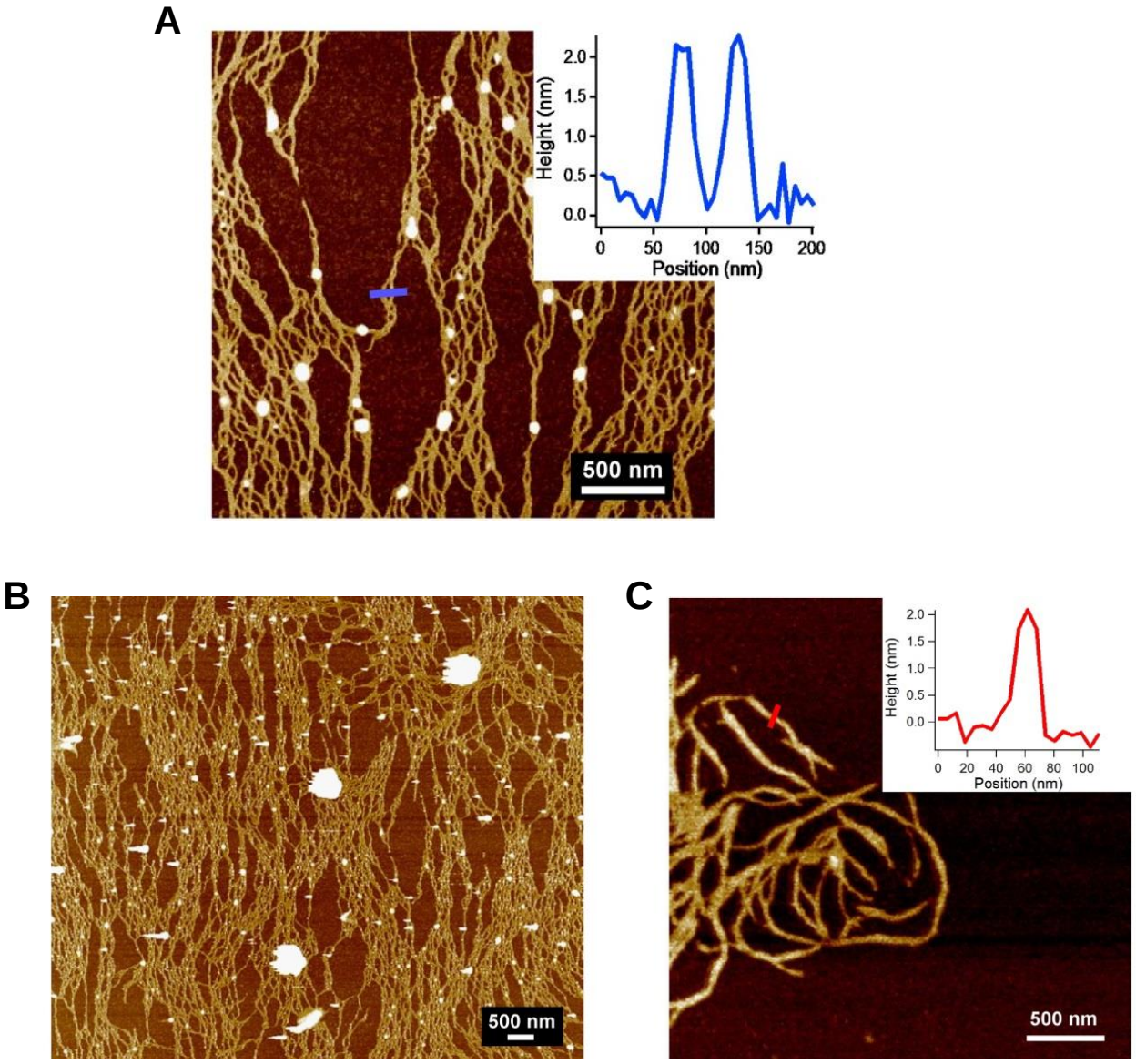


Figure 3.9 AFM topographic images of TAP-CyCo supramolecular structures. (A) Sample containing 15 mM TAP and CyCo6. (B) Same sample in A but with an expanded view. (C) Sample containing 15 mM TAP, 5 mM CyCo4, 10 mM CyCo6 (1:0.33:0.66 ratio of TAP:CyCo4:CyCo6). Inserts shows height profile delineated by the red line in the main panel.

3.3.2 Rheological Properties of TAP-CyCo Hydrogels

Gel mechanical properties were further studied by oscillatory rheology with solutions that were 1.4% by weight in monomer. Hydrogels consisting of either TAP-CyCo6 or TAP-CyCo4-CyCo6 (1:0.33:0.66) exhibited a predominant elastic character over the time scales examined, as evident by a greater storage modulus, G′, than loss modulus, G″ (Figure 3.10). Frequency sweeps were performed at a constant strain value of 1.0%, corresponding to the linear regime as revealed by strain sweeps (Figure 3.10). In addition, strain sweeps indicate that G′ values are similar at pHs near neutrality (ca. +/- 0.5 pH units) for the biomolecular and trimolecular systems, while G″ values are found to vary more markedly with these subtle pH variations and between the systems. More substantial effects were observed when the pH of the solutions was increased to 8.5, as both dynamic moduli decreased by over an order of magnitude. These results indicate that the mechanical properties of the gel phase can be altered with modest pH change.

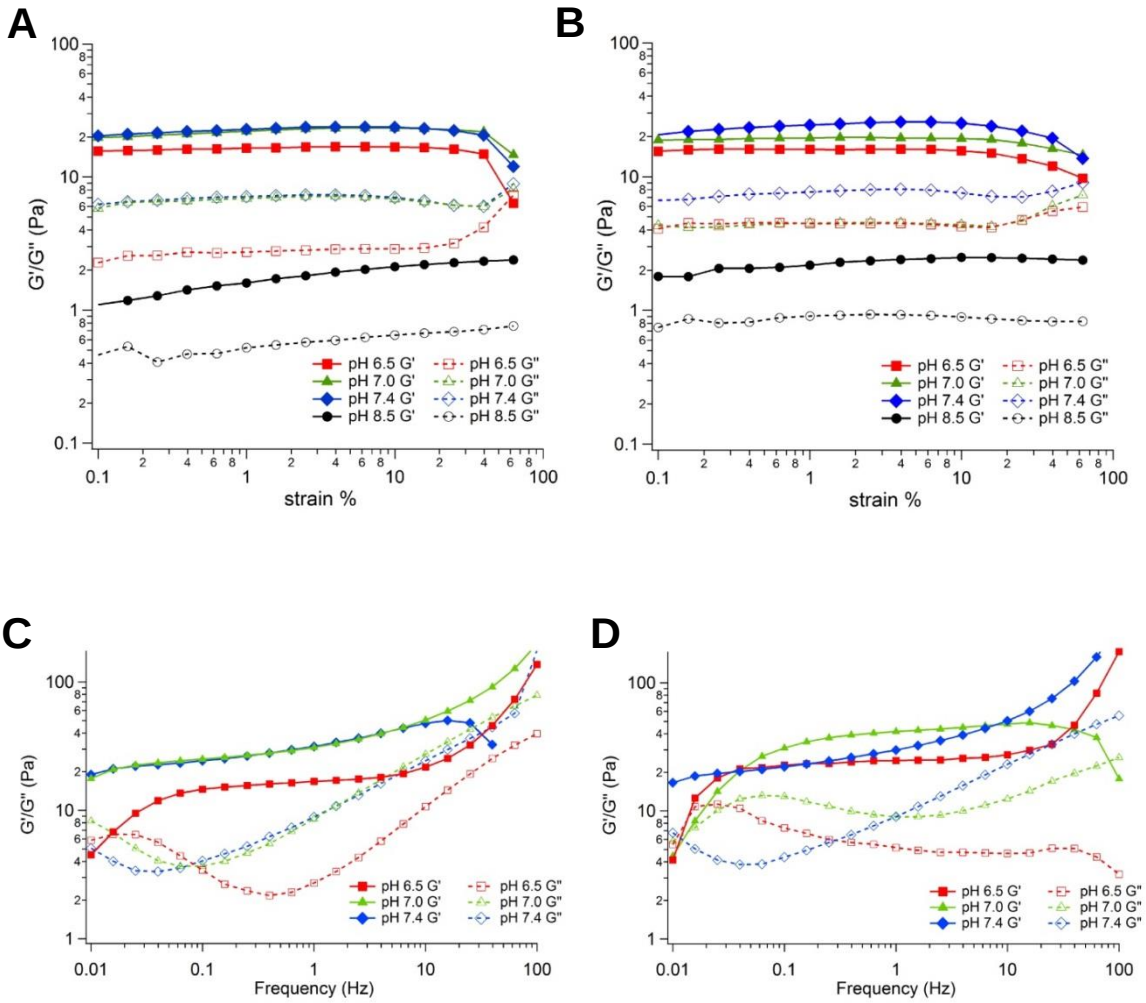


Figure 3.10 Rheological properties of TAP-CyCo hydrogels at various pH. (A) Strain sweeps for gels prepared from TAP and CyCo6 at various pH. (B) Strain sweeps for gels prepared from TAP-CyCo4 and CyCo6 (1:0.66:0.33 ratio) at various pH. (C) Frequency sweeps for gels prepared using TAP-CyCo6 at various pH. (D) Frequency sweeps for gels prepared from TAP, CyCo4 and CyCo6 (1:0.66:0.33 ratio) at various pH. The pHs tested were 6.5, 7.0, 7.4 and 8.5. All solutions were 40 mM in both TAP and CyCo monomer (1.4% by weight in total monomer).

3.3.3 ^{1}H NMR Spectroscopy Reveals a Cooperative Assembly Mechanism

^{1}H NMR spectroscopy indicates the incorporation of these monomers into supramolecular assemblies by exhibiting a concentration-dependent loss of integrated resonance intensity [23]. Specifically, ^{1}H spectra of TAP-CyCo6 solutions below 15 mM in each monomer show CyCo6 resonances with chemical shifts that are identical to CyCo6 alone in solution and with integrated intensities that are directly proportional to concentration (Figure 3.11). In contrast, raising TAP and CyCo6 concentrations above 15 mM results in no further increase of the integrated intensity of CyCo6 resonances, nor are changes observed in ^{1}H chemical shifts. This observation indicates that at the higher concentrations, CyCo6 and TAP are slowly exchanging between their free (unassembled) states and incorporation into large assemblies.

NMR and UV/vis experiments both reveal a cooperative assembly process [24, 25], with a minimal assembly concentration (MAC) for TAP and CyCo6 of 15 mM in each monomer at 20 °C, pH 7. In agreement with the proposed cooperative assembly, NMR and AFM analysis also indicate an absence of intermediate assemblies, i.e., only free monomers and long noncovalent polymers are present. The driving force of this highly cooperative assembly, and thus the formation of long noncovalent polymers, is apparently the hydrophobic effect. The exposure of the two faces of a rosette to water is predicted to have a large positive free energy (ca. 27 kcal/mol) [20, 26]. Moreover, the favourable assembly between TAP and Cy may be enhanced by pK_a matching of these recognition elements, as both

experimental and theoretical studies have indicated that smaller pK_a differences between H-bond donor and acceptor groups leads to stronger H-bonding [27-30].

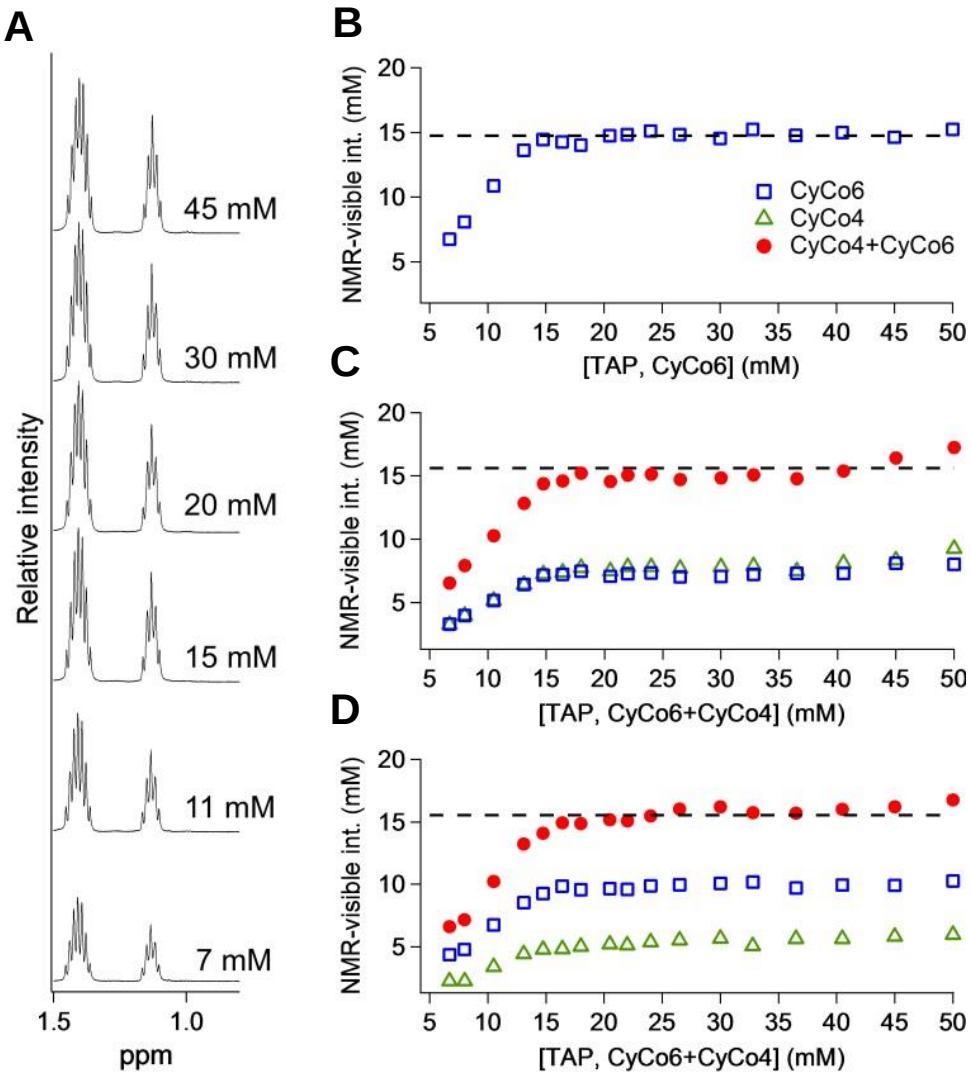


Figure 3.11 ^{1}H NMR of TAP-CyCo assemblies at various concentrations. (A) Representative ^{1}H NMR spectra of the CyCo6 methylene protons for samples containing various concentrations of TAP and CyCo6 (in equimolar ratios). (B-D) Plots of NMR-visible resonance intensity of unassembled CyCo4 and/or CyCo6 vs actual monomer concentrations in (B) 1:1 TAP and CyCo6 solutions, (C) 1:0.5:0.5 solutions of TAP:CyCo4:CyCo6, and (D) 1:0.33:0.66 solutions of TAP:CyCo4:CyCo6. Dashed lines in plots indicate minimal assembly concentrations of the total Cy-containing species.

Trimolecular systems were next investigated by ^{1}H NMR spectroscopy. The analysis shown in Figure 3.11 reveals that the MAC for solutions that contain TAP:CyCo4:CyCo6 in ratios of 1:0.5:0.5 is identical to the MAC measured for the bimolecular TAP-CyCo6 system when the sum of the integrated intensity of CyCo4 and CyCo6 resonances in the trimolecular system are compared to the integrated intensities of CyCo6 resonances in the bimolecular solutions. The same is true for solutions in which the ratio of TAP:CyCo4:CyCo6 is 1:0.33:0.66 (Figure 3.11D). The amount of CyCo4 and CyCo6 incorporated into large assemblies with TAP is simply proportional to the ratio of these monomers present in the samples; equal concentrations of CyCo4 and CyCo6 are incorporated in the 1:0.5:0.5 samples (Figure 3.11C), and a 1:2 ratio of CyCo4 and CyCo6 are incorporated in the 1:0.33:0.66 samples (Figure 3.11D). Additionally, the MAC determined for the mixed systems is the same as that measured for the TAP-CyCo6 system at pH 7 for temperatures ranging from at least 5 to 35 ºC (Figure 3.12). These observations confirm that the mixed trimolecular system assembles with the same propensity as the TAP-CyCo6 system and demonstrates that formation of soluble co-assemblies with CyCo4 and CyCo6 are insensitive to the differences in the lengths of their carboxylic acid tails.

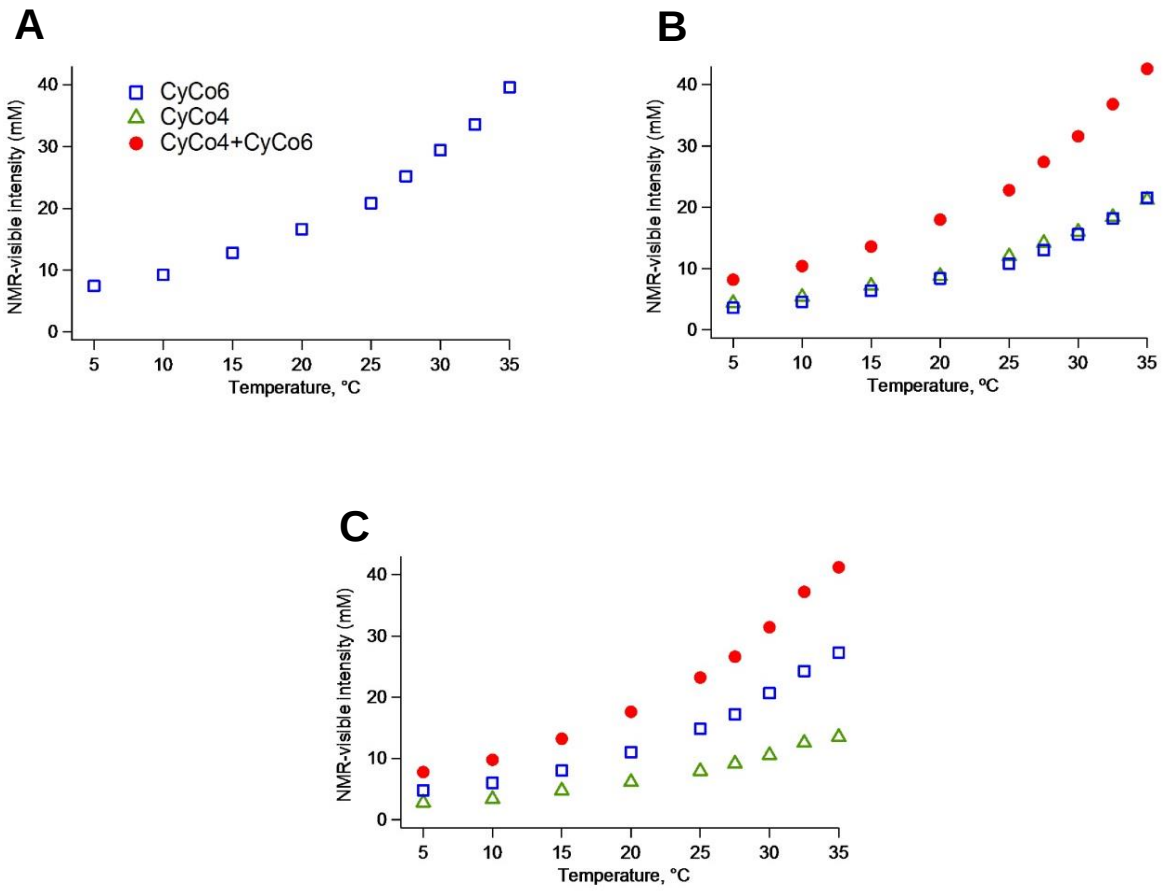


Figure 3.12 Plots of the apparent solution-phase concentrations (equivalent to the MAC) of TAP-CyCo assemblies vs temperature. All solutions contained 40 mM TAP and 40 mM CyCo4+CyCo6 at (A) 1:0:1, (B) 1:0.5:0.5 or (C) 1:0.33:0.66 molar ratios of TAP:CyCo4:CyCo6. Solutions contained 40 mM TAP and 40 mM CyCo4+CyCo6.

3.3.4 Precipitation of Assemblies

The precipitate of a 1:0.5:0.5 ratio TAP:CyCo4:CyCo6 sample was found to be enriched in CyCo4 (Figure 3.13), which is likely the result of monomer rearrangement producing regions enriched in TAP-CyCo4 assemblies that precipitate due to the greater propensity of this particular assembly to aggregate. This proposed mechanism of precipitation and observed precipitate enrichment in CyCo4 is consistent with a recent report by Meijer and coworkers, which showed that rearrangement of monomers within and between noncovalent polymers in water can require on the order of hours [31].

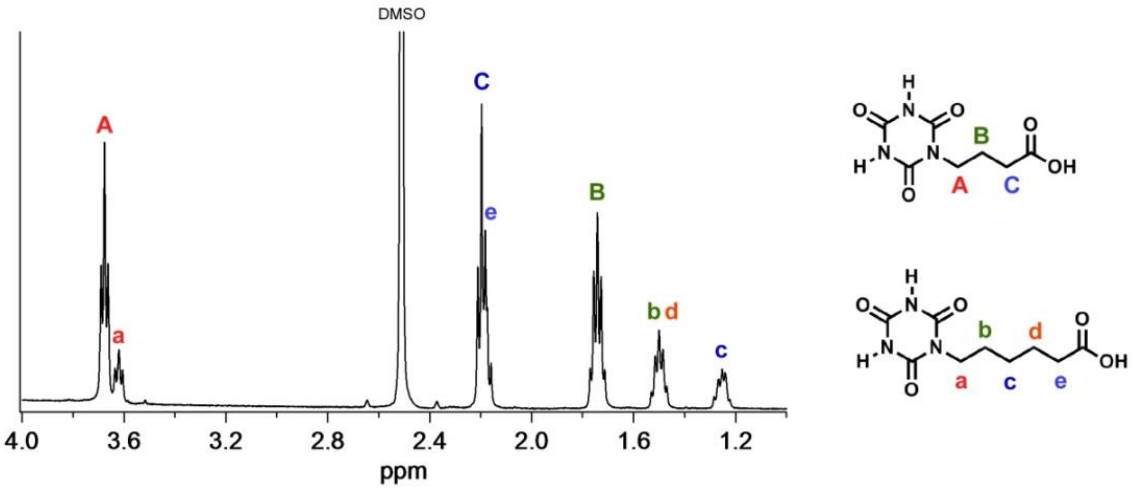


Figure 3.13 ^{1}H NMR spectrum of precipitated material collected from a solution originally containing a 1:0.5:0.5 ratio of TAP:CyCo4:CyCo6. TAP and CyCo4+CyCo6 were originally 40 mM in 200 mM sodium phosphate buffer, pH 7. Precipitation from the gel phase began after two hours, resulting in complete gel collapse over the course of 24 hours. The NMR spectrum of the precipitate dissolved in DMSO reveals precipitate enrichment in CyCo4, with a 4:1 ratio of CyCo4:CyCo6. Resonance assignments indicated on the CyCo4 and CyCo6 chemical structure were used to determine the ratio of both molecules in the isolated precipitate.

Precipitation of TAP-CyCo4 and TAP-CyCo6 assemblies below pH 7 likely results from intermolecular hydrogen bonding of the neutral carboxylic acid groups that bundle the linear assemblies into large aggregates, as has been seen with other supramolecular polymerization systems that contain monomers with carboxylic acid groups [11]. Cooperative interactions of the mixed monomers are highly synergistic, enabling both the CyCo4 monomer to form long-lived gels and extending the lifetime of CyCo6 assemblies under unfavorable conditions (i.e., low pH). It has been previously shown that substantial modifications in the peripheral groups of recognition elements that also form rosette assemblies can result in the adoption of alternative assembly modes [32, 33]. However, less substantial changes are known to lead to soluble rosette nanotubes or insoluble tightly packed rosette assemblies [34-36]. As demonstrated with this study, the addition of a structurally similar co-monomer allows the rosette assembly unit to be maintained while simultaneously leading to large differences in assembly solubility due to the frustration of regular packing (and thus crystallization) between the assemblies.

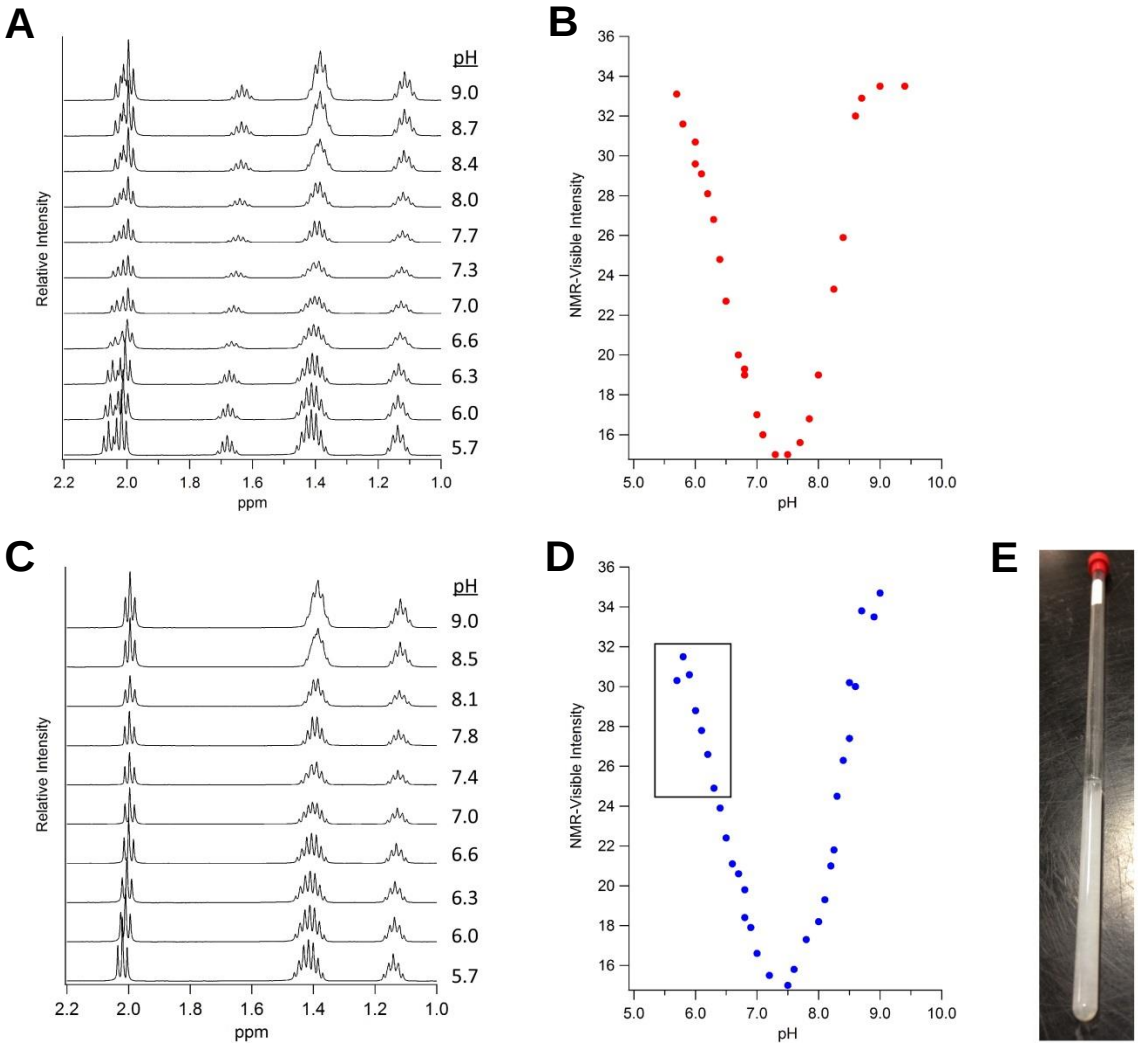


Figure 3.14 1H-NMR analysis of the effect of pH on TAP-CyCo assemblies. (A) Representative spectra of a solution containing a 1:0.33:0.66 ratio of TAP:CyCo4:CyCo6, at 35 mM in TAP and CyCo4+CyCo6. (B) Plot of the apparent solution-phase concentration as determined from the NMR experiment shown in (A) of CyCo6 + CyCo4 vs pH, which provides the MAC as a function of pH. (C) Representative spectra of a solution containing a 1:1 ratio of TAP:CyCo6 originally at 40 mM in each monomer. (D) Plot of the apparent solution-phase concentration as determined from NMR experiment shown in C of CyCo6 vs pH. Note that the box in left corner of plot indicates samples that contained TAP-CyCo6 precipitate. (E) NMR tube containing a solution with a white precipitate of TAP and CyCo6 at pH 6.5 from the titration experiment.

3.3.5 Ultrasensitive pH Control of TAP-CyCo Assemblies

The pH dependent assembly of both bimolecular (TAP-CyCo6) and trimolecular (TAP-CyCo6-CyCo4) supramolecular polymers were investigated by ^{1}H NMR spectroscopy. Any deviation in pH away from 7.4 corresponded to a shift in equilibrium of the supramolecular polymer away from assembly (Figure 3.14 and 3.15). Complete disassembly occurs when the pH is below 5.7 or above pH 9.0. As anticipated, precipitation was evident when the bimolecular system was evaluated at any pH below 6.5 (Figure 3.14). However, for the three-component system, no precipitation was observed over this entire pH range.

The behavior of the trimolecular system quantitatively illustrates the high level of structural control that can be achieved with pK_a-matched monomers. Specifically, a comparison of the experimental data to the theoretical limit for a multimolecular assembly as a function of pH demonstrates that the TAP-CyCo system has the theoretically maximum sensitivity to change in pH that can be obtained for a supramolecular polymerization system (Figure 3.15). An observation that is consistent with the TAP-CyCo assemblies being the most pH responsive supramolecular polymers ever reported. We note that our observation of the TAP-CyCo system being maximally assembled at pH 7.4 indicates pK_a-matched monomers at this pH, which is about 0.5 pH units above the reported pK_as of TAP and Cy. Consistent with this observation, the apparent pK_as of TAP, CyCo6 and CyCo4 were individually determined to each be around pH 7.4 in the buffer conditions of our experiments.

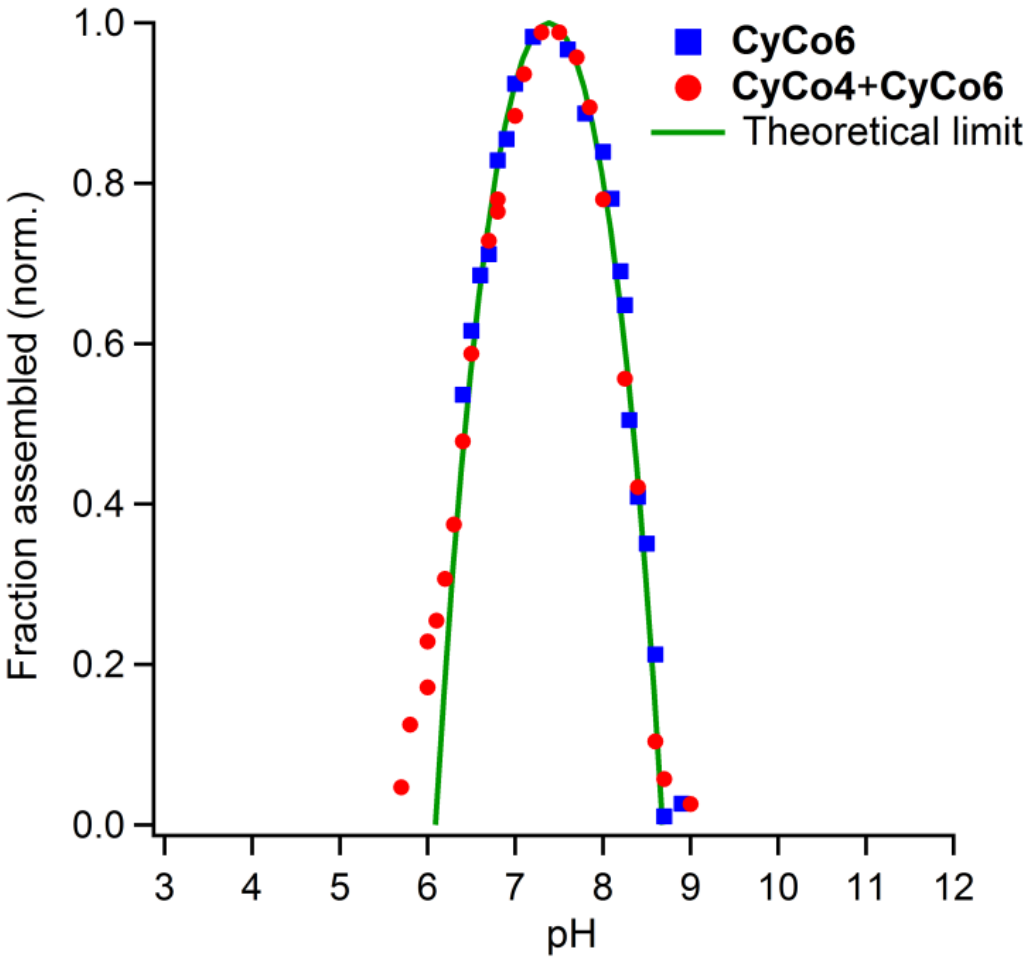


Figure 3.15 Plot of the normalized fraction of CyCo/TAP monomer assembled in solutions that were 35 mM in TAP and CyCo monomer based on the ^{1}H NMR visible concentration of unassembled CyCo6 (in 1:1 TAP:CyCo6 solutions) (blue) and CyCo4+CyCo6 (in 1:0.33:0.66 TAP:CyCo4:CyCo6 solutions) (red) as a function of pH. The theoretical limit governing supramolecular assembly for a pK_a-matched system as a function of pH is shown as a green line. See caption of Figure 3.1 and section 3.6 for additional model details and derivation. Note that below pH 6.3 the solution containing only TAP and CyCo6 immediately formed a precipitate, so data points for mixtures below this pH could not be obtained.

3.4 Conclusion

Presented here is a new design strategy for the formation of supramolecular structures that are optimally responsive to changes in pH by utilizing pK_a matched complementary monomeric recognition elements, acidic and basic monomers. Systems that are pK_a matched enable dual pH control of the assembly, which is most favoured in a solution in which the pH equals the pK_a of the monomers. The TAP-Cy assembly motif is exceptional because both recognition elements have pK_as near neutrality, making systems that employ these complimentary heterocycles attractive for the development of materials useful for biological applications, as small pH changes near physiological pH will elicit an appreciable and predicable response [13, 37, 38]. For example, materials that can undergo a structural transition upon encountering slightly acidic conditions, as found in the extracellular microenvironment of some tumours (pH 6.3-7.2), could find applications in drug delivery and diagnostics [13, 39]. A significant observation of this study is that heterogeneity within the supramolecular polymers stabilizes the hydrogel phase against precipitation. The ability of mixed assemblies formed by multiple monomers with the same recognition elements but different side chains that suppress the interactions that cause precipitation further supports this approach as a general principle for the design of well-behaved, self-assembling noncovalent polymer systems.

Finally, this study may provide insight into a potential mechanism that allowed for replication before the advent of macromolecular catalysts (i.e., enzymes). As discussed in the preceding chapters the emergence of polymers capable of storing information (i.e., genetic polymers) is widely believed to have been a critical step in the origin of life, a prerequisite for the hypothetical 'RNA world' [40]. And the first genetic

polymers to emerge may have been ancestral forms of the contemporary nucleic acids (i.e., RNA-like) that were of different chemical structure (i.e., with different bases and a different backbone) [41-43]. If TAP and Cy nucleosides were incorporated into early informational polymers, then the effects of pK_a matching reported here may provide an advantage for non-enzymatic replication before the advent of enzymes. In particular, small changes in pH would have altered the affinity of the recognition units and thereby promoted strand separation. This property could have aided early replication cycles by having strand separation be regularly favored by geochemically driven changes in pH. Furthermore, this proposal is attractive as it does not necessitate high temperatures for strand separation that could result in polymer degradation. For example, RNA duplexes composed of moderately sized strands (50 nts or greater) typically require temperature of 90 ºC or higher for separation (which greatly promotes backbone hydrolysis [44].

3.5 Calculations

3.5.1 Derivation of Equation for pH Sensitivity

Consider a system of two molecules, an acid **AH** and a base **B**, that assemble in aqueous solution. The deprotonated (i.e., negatively charged) form of **AH** is designated as $\mathbf{A}^{\ominus}$, and the protonated (i.e., positively charged) form of **B** is designated as $\mathbf{BH}^{\oplus}$.

If the neutral forms of these molecules assemble with a 1:1 stoichiometry, then we can define a solubility product for these monomers that is the product of the concentrations of the free (uncharged) monomers that coexist in solution with the supramolecular assembly. That is,

$$K_{sp} = [\mathbf{AH}][\mathbf{B}] \qquad \text{(Eq. 1)}$$

We can write the Henderson-Hasselbalch equations for both monomers as

$$\mathrm{pH} = \mathrm{p}K_{\mathrm{AH}} + \log\frac{[\mathbf{A}^{\ominus}]}{[\mathbf{AH}]} \quad \text{and} \quad \mathrm{pH} = \mathrm{p}K_{\mathrm{BH}^{\oplus}} + \log\frac{[\mathbf{B}]}{[\mathbf{BH}^{\oplus}]},$$

where $\mathrm{p}K_{\mathrm{AH}}$ is the pK_a of **AH** and $\mathrm{p}K_{\mathrm{BH}^{\oplus}}$ is the pK_a of the conjugate acid of **B**. Using these equations, the concentrations of the ionized forms of these molecules can be related to the concentrations of their neutral forms,

$$[\mathbf{A}^{\ominus}] = [\mathbf{AH}] \times 10^{\mathrm{pH}-\mathrm{p}K_{\mathrm{AH}}} \quad \text{(Eq. 2)},$$

$$[\mathbf{BH}^{\oplus}] = [\mathbf{B}] \times 10^{\mathrm{p}K_{\mathrm{BH}^{\oplus}}-\mathrm{pH}} \text{(Eq. 3)}.$$

Here, to simplify our derivation, we consider only samples with equal molar amounts of each monomer. Because the supramolecular assemblies and the sample as a whole both have a 1:1 stoichiometry of **AH** and **B**, the concentrations of the two monomers free in solution will also be equal. That is, [**AH**] + [$\mathbf{A}^{\ominus}$] = [**B**] + [$\mathbf{BH}^{\oplus}$]. Using Eq. 2 and Eq. 3, we can rewrite this equality as:

$$[\mathbf{AH}](1 + 10^{\mathrm{pH}-\mathrm{p}K_{\mathrm{AH}}}) = [\mathbf{B}](1 + 10^{\mathrm{p}K_{\mathrm{BH}^{\oplus}}-\mathrm{pH}}) \qquad \text{(Eq. 4)}$$

Using Eq. 1 to eliminate [**B**] from Eq. 4, and solving for [**AH**] we obtain,

$$[\mathbf{AH}] = \sqrt{\frac{K_{\mathrm{sp}}(1+10^{\mathrm{p}K_{\mathrm{BH}^{\oplus}}-\mathrm{pH}})}{(1+10^{\mathrm{pH}-\mathrm{p}K_{\mathrm{AH}}})}}. \qquad \text{(Eq. 5)}$$

The total concentration of the neutral and ionized species of **AH** in solution can then be written as,

$$[\mathbf{AH}] + [\mathbf{A}^{\ominus}] = \sqrt{K_{\mathrm{sp}}(1 + 10^{\mathrm{p}K_{\mathrm{BH}^{\oplus}}-\mathrm{pH}})(1 + 10^{\mathrm{pH}-\mathrm{p}K_{\mathrm{AH}}})}. \qquad \text{(Eq. 6)}$$

We define the Normalized Fraction Assembled, or NFA, as the relative fraction of monomers assembled at a particular pH as compared to the fraction of monomers

assembled at the pH of maximum assembly, which is midway between pK_{AH} and $pK_{BH^{\oplus}}$ (pK_{ave}). That is,

$$\mathrm{NFA} = \frac{\mathbf{A}_{tot} - \sqrt{K_{sp}(1+10^{pK_{BH^{\oplus}}-pH})(1+10^{pH-pK_{AH}})}}{\mathbf{A}_{tot} - \sqrt{K_{sp}}(1+10^{pK_{ave}-pK_{AH}})}, \qquad \text{(Eq. 7)}$$

where $\mathbf{A}_{tot}$ is equal to the *total* concentration of **AH** in the sample, in neutral and ionized forms, and as free monomers and in supramolecular assemblies. The positive values of Eq. 7, ranging from 0 to 1, represent physically meaningful solutions to the NFA; negative values indicate samples with completely unassembled monomers.

We note that a similar equation is obtained if the ionized species of **AH** and **B** form the supramolecular assembly. In which case $K_{sp} = [\mathbf{A}^{\ominus}][\mathbf{BH}^{\oplus}]$, and

$$\mathbf{NFA} = \frac{\mathbf{A}_{tot} - \sqrt{K_{sp}\left(1+10^{pH-pK_{BH^{\oplus}}}\right)\left(1+10^{pK_{AH}-pH}\right)}}{\mathbf{A}_{tot} - \sqrt{K_{sp}}(1+10^{pK_{AH}-pK_{ave}})}.$$

3.5.2 Examples Calculations

For the experimental data presented in the main text, CyCo6+CyCo4 is AH; TAP is B. The pK_{AH} of CyCo6 and CyCo4 are equal to the $pK_{BH^{\oplus}}$ of TAP, which is set here to 7.4. Experiments were carried out at pH 7.4 with total concentration TAP of 35 mM and total CyCo6+CyCo4 concentration of 35 mM, exhibiting a free monomer concentration of 15 mM in TAP and 15 mM in CyCo6+CyCo4. The concentrations of the *neutral* species of **AH** and **B** are therefore each 7.5 mM at the pH of maximum assembly. Thus, $\mathbf{A}_{tot}$ = 35 mM, K_{sp} = (7.5 mM)^2, $pK_{ave} = pK_{AH}$, and Eq. 7 reduces to:

$$\mathbf{NFA} = \frac{35 - 7.5\sqrt{(1+10^{7.4-pH})(1+10^{pH-7.4})}}{20},$$

which gives the green curve shown in Figures 3.1 and 3.15 when plotted versus pH. If **AH** and **B** are not pK_a matched (i.e., $pK_{AH} < pK_{BH\oplus}$), but the neutral monomers assemble with the same K_{sp}, then the **NFA** of solutions containing 35 mM in each monomer (neutral and ionized, free and assembled), is given by:

$$\mathbf{NFA} = \frac{35 - 7.5\sqrt{(1+10^{pK_{BH\oplus}-pH})(1+10^{pH-pK_{AH}})}}{35 - 7.5(1+10^{pK_{ave}-pK_{AH}})}.$$

For example, if pK_{ave} of **AH** and **B** is still 7.4, but their ΔpK_a is 5 (i.e., pK_{AH} = 4.9 and $pK_{BH\oplus}$ = 9.9), then the **NFA** curve is given by:

$$\mathbf{NFA} = \frac{35 - 7.5\sqrt{(1+10^{9.9-pH})(1+10^{pH-4.9})}}{35 - 7.5(1+10^{2.5})},$$

which is the red curve shown in Figure 3.1.

3.7 References

1. Wojtecki, R.J., M.A. Meador, and S.J. Rowan, *Using the dynamic bond to access macroscopically responsive structurally dynamic polymers.* Nature Mater, 2011. **10**: p. 14-27.

2. Aida, T., E.W. Meijer, and S.I. Stupp, *Functional supramolecular polymers.* Science, 2012. **335**: p. 813-817.

3. Hartgerink, J.D., E. Beniash, and S.I. Stupp, *Self-assembly and mineralization of peptide-amphiphile nanofibers.* Science, 2001. **294**: p. 1684-1688.

4. Hirst, A.R., et al., *High-tech applications of self-assembling supramolecular nanostructured gel-phase materials: from regenerative medicine to electronic devices.* Angew Chem Int Ed Engl, 2008. **47**: p. 8002-8018.

5. Besenius, P., et al., *Controlling the growth and shape of chiral supramolecular polymers in water.* Proc Natl Acad Sci USA, 2010. **107**: p. 17888-17893.

6. Fenske, T., et al., *Advances in Switchable Supramolecular Nanoassemblies.* Chem Eur J, 2012. **18**: p. 738-755.

7. Buerkle, L.E. and S.J. Rowan, *Supramolecular gels formed from multi-component low molecular weight species.* Chem Soc Rev, 2012. **41**: p. 6089-6102.

8. Segarra-Maset, M.D., et al., *Control of molecular gelation by chemical stimuli.* Chem Soc Rev, 2013. **42**: p. 7086-98.

9. Frisch, H. and P. Besenius, *pH-Switchable Self-Assembled Materials.* Macromol Rapid Commun, 2014. DOI: 10.1002/marc.201400623

10. Frisch, H., et al., *pH-Regulated Selectivity in Supramolecular Polymerizations: Switching between Co- and Homopolymers.* Chemistry, 2015. **21**: p. 3304-3309.

11. Tang, C., et al., *Fmoc-diphenylalanine self-assembly mechanism induces apparent pKa shifts.* Langmuir, 2009. **25**: p. 9447-9453.

12. Adams, D.J., et al., *The delicate balance between gelation and crystallisation: structural and computational investigations.* Soft Matter, 2010. **6**: p. 4144.

13. Ghosh, A., et al., *Fine-tuning the pH trigger of self-assembly.* J Am Chem Soc, 2012. **134**: p. 3647-3650.

14. Frisch, H., et al., *pH-Switchable ampholytic supramolecular copolymers.* Angew Chem Int Ed Engl, 2013. **52**: p. 10097-10101.

15. Fenske, T., et al., *Cooperative Self-Assembly of Discoid Dimers: Hierarchical Formation of Nanostructures with a pH Switch.* J Am Chem Soc, 2013: p. 8342–8349.

16. Fleischer, M. and C. Schmuck, *Transforming polyethylenimine into a pH-switchable hydrogel by additional supramolecular interactions.* Chem Commun, 2014. **50**: p. 10464-10467.

17. Seto, C.T. and G.M. Whitesides, *Self-assembly based on the cyanuric acid melamine lattice.* J Am Chem Soc, 1990. **112**: p. 6409-6411.

18. Lehn, J.-M., et al., *Molecular recognition directed self-assembly of ordered supramolecular strands by cocrystallization of complementary molecular components.* J Chem Soc, Chem Commun, 1990: p. 479.

19. Ma, M. and D. Bong, *Determinants of cyanuric acid and melamine assembly in water.* Langmuir, 2011. **27**: p. 8841-8853.

20. Cafferty, B.J., et al., *Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues.* J Am Chem Soc, 2013. **135**: p. 2447-2450.

21. Hager, K., A. Franz, and A. Hirsch, *Self-assembly of chiral depsipeptide dendrimers.* Chem Eur J, 2006. **12**: p. 2663-2679.

22. Buerkle, L.E., et al., *Tailoring the properties of guanosine-based supramolecular hydrogels.* Langmuir, 2009. **25**: p. 8833-8840.

23. Duncan, D.C. and D.G. Whitten, *1H NMR Investigation of the Composition, Structure, and Dynamics of Cholesterol−Stilbene Tethered Dyad Organogels.* Langmuir, 2000. **16**: p. 6445-6452.

24. Hirst, A.R., et al., *Low-molecular-weight gelators: elucidating the principles of gelation based on gelator solubility and a cooperative self-assembly model.* J Am Chem Soc, 2008. **130**: p. 9113-9121.

25. De Greef, T.F., et al., *Supramolecular polymerization.* Chem. Rev., 2009. **109**: p. 5687-5754.

26. Chandler, D., *Interfaces and the driving force of hydrophobic assembly.* Nature, 2005. **437**: p. 640-647.

27. Perrin, C.L. and J.B. Nielson, *"Strong" hydrogen bonds in chemistry and biology.* Annu Rev Phys Chem, 1997. **48**: p. 511-544.

28. Chen, J., et al., *Short, Strong Hydrogen Bonds in the Gas Phase and in Solution: Theoretical Exploration of pKaMatching and Environmental Effects on the Strengths of Hydrogen Bonds and Their Potential Roles in Enzymatic Catalysis.* J Org Chem, 1998. **63**: p. 4611-4619.

29. Acharya, P., et al., *Measurement of nucleobase pK_a values in model mononucleotides shows RNA-RNA duplexes to be more stable than DNA-DNA duplexes.* J Am Chem Soc, 2004. **126**: p. 2862-2869.

30. Engelhart, A.E., T.H. Morton, and N.V. Hud, *Evidence of strong hydrogen bonding by 8-amino-guanine.* Chem Commun, 2009: p. 647-649.

31. Albertazzi, L., et al., *Spatiotemporal control and superselectivity in supramolecular polymers using multivalency.* Proc Natl Acad Sci, 2013. **110**: p. 12203-12208.

32. Whitesides, G.M., et al., *Noncovalent Synthesis: Using Physical-Organic Chemistry To Make Aggregates.* Acc Chem Res, 1995. **28**: p. 37-44.

33. Bielejewska, A.G., et al., *Thermodynamic Stabilities of Linear and Crinkled Tapes and Cyclic Rosettes in Melamine−Cyanurate Assemblies: A Model Description.* J Am Chem Soc, 2001. **123**: p. 7518-7533.

34. Marsh, A., M. Silvestri, and J.-M. Lehn, *Self-complementary hydrogen bonding heterocycles designed for the enforced self-assembly into supramolecular macrocycles.* Chem Commun, 1996: p. 1527.

35. Mascal, M., et al., *The G−C DNA Base Hybrid: Synthesis, Self-Organization and Structural Analysis.* J Org Chem, 1999. **64**: p. 8479-8484.

36. Durmus, A., et al., *Synthesis of N-substituted Pyrido[4,3-d]pyrimidines for the Large-Scale Production of Self-Assembled Rosettes and Nanotubes.* J Org Chem, 2013. **78**: p. 11421-11426.

37. Schmaljohann, D., *Thermo- and pH-responsive polymers in drug delivery.* Adv Drug Deliv Rev, 2006. **58**: p. 1655-1670.

38. Stuart, M.A., et al., *Emerging applications of stimuli-responsive polymer materials.* Nature Mater, 2010. **9**: p. 101-113.

39. Mura, S., J. Nicolas, and P. Couvreur, *Stimuli-responsive nanocarriers for drug delivery.* Nature Mater, 2013. **12**: p. 991-1003.

40. Joyce, G.F., *The antiquity of RNA-based evolution.* Nature, 2002. **418**: p. 214-221.

41. Joyce, G.F., et al., *The case for an ancestral genetic system involving simple analogs of the nucleotides.* Proc Natl Acad Sci, 1987. **84**: p. 4398-402.

42. Engelhart, A.E. and N.V. Hud, *Primitive genetic polymers.* Cold Spring Harbor Perspectives in Biology, 2010. **2**. DOI 10.1101/cshperspect.a002196.

43. Hud, N.V., et al., *The origin of RNA and 'My Grandfather's Axe'.* Chem Biol, 2013. **20**: p. 466-474.

44. Szostak, J.W., *The eightfold path to non-enzymatic RNA replication.* J Sys Chem, 2012. **3**: p. 2-15.

CHAPTER 4

EVALUATION OF BIMOLECULAR ROSETTE ASSEMBLIES IN WATER USING MOLECULAR SCREENING AND MONOMER SUBSTITUTION[d]

4.1 Introduction

Monomeric compounds that self-assemble in water to form non-covalent supramolecular polymers and hydrogels are of current interest because of their potential applications in biomedical technologies [1-5]. While many of these systems have been discovered serendipitously or by screening libraries of related compounds, continued development of functional water-based supramolecular polymers, and hydrogels based on these systems, will require a thorough understanding of the interactions that can be employed [6-8]. Typically, supramolecular polymers rely on hydrogen bonding, dispersion forces, ion pairing, metal-ligand coordination and solvophobic interactions that are incorporated, or influenced, through monomer design and can be manipulated by environmental changes [9-12]. An additional level of control may be realized by designing multimolecular assemblies, or copolymers, which can enable a more facile tailoring of a system's properties [13, 14]. Regardless, the discovery of new aqueous systems and their ability to function in complex chemical environments (e.g., in vivo) will ultimately require further understanding of the factors enabling, or limiting, supramolecular assembly in water.

d. Work presented in this chapter was performed in collaboration with Dr. Chao Li.

Synthetic assembly systems that utilize, or were inspired by, nucleic acid base-pairing interactions have been widely employed in nanotechnology [15-17]. Monomeric assembly (i.e., untethered recognition elements) of heterocycles that engage in Watson-Crick-like interactions are limited to a small number of heterocycles including derivized guanosine and hexameric rosettes [18]. Extensive investigations of rosette assembly motifs, both monomolecular [19-25] (e.g., the Lehn-Mascal G-C base) and biomolecular [26-33] (e.g., barbituric acid and triaminopyrimidine or melamine and cyanuric acid), have been predominantly performed in organic solvents. Assembly of hydrogen bonded rosettes in aqueous media has been largely neglected in part due to the supposition that hydrogen bonding within the assembly cannot compete with hydrogen bonding between the monomers and solvent.

A few studies have been carried out with the self-complementary Janus-type monomolecular assemblies in water. Specifically, Fenniri and coworkers have reported the assembly of self-complementary bi- and tricyclic heterocycles into rosettes possessing 18-hydrogen bonds, which stack to form long supramolecular polymers (Figure 4.1) [34-36]. Several water-soluble peripheral groups have been substituted on the heterocycle (including amino acids and crown-ethers) and the heterocycles could be tethered with charged linkers and still maintain the ability to assemble in water [35, 37, 38]. Additionally, they showed that the bicyclic ring system originally employed for Janus assembly systems could be expanded into a tricyclic system by the inclusion of an internal pyridine ring (Figure 4.1) [39]. Increasing stacking interactions would be expected to lead to more favorable assembly. Thus it would be expected that the

formation of rosette-based supramolecular polymers from heterocycles with larger ring systems would become more favorable.

As discussed in the previous chapters, water-soluble supramolecular polymers form from stacked bimolecular rosettes containing monocyclic triaminopyrimidine (TAP) and cyanuric acid (Cy) that, like Fenirri's Janus-type system, are also held together by 18-hydrogen bonds but possess a considerably smaller surface area [40-42]. In our previous studies of these systems, basic design principles were established that enable TAP and Cy to form water soluble assemblies (i.e., modification of one of the heterocycles with a hydrophilic peripheral group to block formation of planar sheets and increase solubility of the supramolecular polymers), however no evaluation of the heterocyclic components has been performed. In this chapter, we examine the ability of monomers to form rosette-based supramolecular polymers in water by screening heterocycles that are closely related to TAP and Cy. The importance of ring size and hydrogen bonding is systematically evaluated by testing heterocycles with the purine or pyrimidine ring system for incorporation into TAP-Cy supramolecular polymers. We demonstrate that the purine analogs of TAP, 2,6-diaminopurine (DAP) and 2,4-diaminopyrimidine, can substitute for TAP and assemble with a Cy substituted monomer to form supramolecular polymers and hydrogels. Our experiments reveal the importance of maintaining 18 hydrogen bonds and the matching of hydrophobic surfaces in the design of aqueous rosette assemblies, and suggest that the number of heterocycles that can form these assemblies is very limited, a finding that has implications for the development of water-based supramolecular polymers and materials; as well as into investigations related to uncovering the nucleobase of proto-RNA.

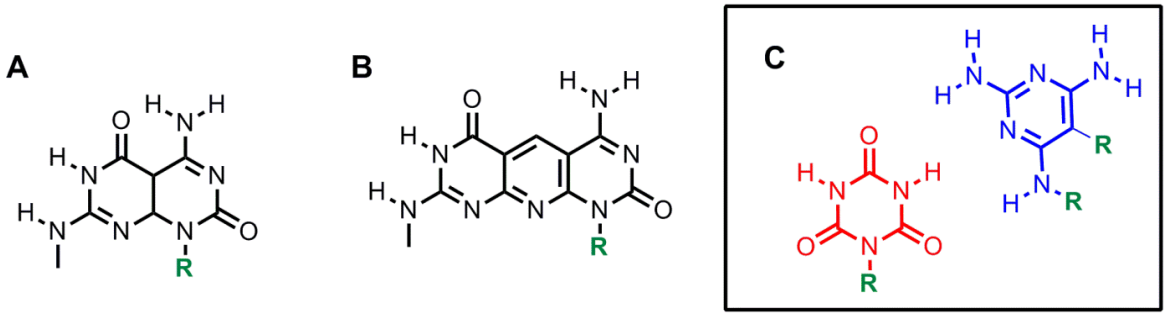


Figure 4.1 Heterocycles that have been reported to assemble into hexameric rosettes as uncoupled monomers in water. (A) Bicyclic and (B) tricyclic self-complementary Janus-type G-C hybrid molecules demonstrated by Fenniri and coworkers to form monomolecular assemblies. (C) TAP and Cy monomers demonstrated by Hud and coworkers to form biomolecules assemblies. R represents positions that have been modified with hydrophilic exocyclic groups.

4.2 Experimental Procedures

4.2.1 Materials and Synthesis

CyCo4, CyCo6 [43] and 2-hydroxypurine [44] were synthesized according to previous reports. Other pyrimidines and purines were commercial available, and used as received.

CyCo4): ^{1}H NMR (500 MHz, DMSO-d_6): δ = 1.749 (m, 2H; CH_2), 2.243 (t, J = 7.5 Hz, 2H; CH_2CO), 3.675 (t, J = 7 Hz, 2H; CH_2N), 11.63 ppm (brm, NH); ^{13}C NMR (125 MHz, DMSO-d_6): δ = 23.23, 31.42, 40.47, 149.15, 150.44, 174.47 ppm; MS (m/z): $[M-H]^-$ 213.8.

CyCo6: ^{1}H NMR (300 MHz, DMSO-d_6): δ = 1.32 (m, 2H; CH_2), 1.64 (m, 4H; CH_2), 2.30 (t, ^{3}J=7.7 Hz, 2H; CH_2CO), 3.71 (t, ^{3}J=7.7 Hz, 2H; CH_2N), 11.63 ppm (brm, 2H; NH);

^{13}C NMR (75 MHz, DMSO-d_6): δ = 24.6, 25.8, 27.3, 33.8, 39.1, 149.0, 150.3, 174.6 ppm; ESI-MS (*m/z*): 241.9 (M^-)

2-hydroxypurine: ^{1}H NMR (300 MHz, DMSO-d_6,): δ = 8.63 (s, 1H), 8.92 ppm (s, 1H); ^{13}C NMR (75 MHz, DMSO-d_6): δ = 118.2, 142.0, 148.9, 155.4, 164.3 ppm; ESI-MS (*m/z*): 137.1 (M+H).

4.2.2 Sample Preparation

Stock solutions of xanthine contained 5 mM molecule, and were adjusted to pH 10 with 1 M NaOH to maintain solubility. The other heterocycles were dissolved in unbuffered water with the concentration of stock solutions depending on the solubility of each compound. Stock solutions of TAP and CyCo6 contained 200 mM and 100 mM molecules, respectively. All assembly solutions were analyzed within 5 h of preparation.

4.2.3 Instrumental Methods

Spectroscopic analyses of the assemblies were performed using UV-vis and ^{1}H NMR. UV-vis analysis was carried out on an Agilent 8453 spectrophotometer equipped with an 89090A temperature controller. To maintain an optical density below 1.3, cells of different path lengths (0.1 and 0.01 mm) were used. ^{1}H NMR spectra were collected on a Bruker DRX-500 500 MHz NMR and were the sum of 32 transients in D_2O. All samples investigated by NMR were D_2O exchanged before addition of the internal standard trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP) at 0.56 mM. TSP did not show any indication of interacting with the assemblies. Determination of CyCo6-DAP precipitate composition was performed in DMSO-d_6.

AFM imaging was performed on freshly cleaved mica that was pre-activated by adding 40 μL of 20 mM $MgCl_2$ with 40 min incubation. The mica was rinsed extensively with water and dried under N_2 (*g*). A 10-μL sample of the assembly solution was spin coated for 30 sec at 4000 rpm and dried at room temperature overnight. Samples were incubated on ice to promote assembly and were applied to the mica surface while still cold. AFM imaging was performed with a Nanoscope IIIa (Digital Instruments) in tapping mode in air, using Si tips (Vistaprobes, 48 N/m).

Bimolecular assemblies were formed by adding CyCo6 to a solution containing an equivalent amount of TAP or DAP and 200 mM sodium phosphate buffered at pH 7. Unless otherwise noted, trimolecular assembly solutions were prepared containing 30 mM TAP, 30 mM CyCo6 and 5 mM of the heterocycle (1-18) being evaluated with 200 mM sodium phosphate buffer. DAP-CyCo4 assemblies were formed by first heating the solutions until all components were in dissolved, followed by cooling on ice.

POM images were obtained using a Leica DMRX optical microscope equipped with a rotatable polarizer and analyzer and a Nikon D300 digital SLR camera.

4.3 Results and Discussion

4.3.1 Screening for Incorporation of Heterocycles into Rosette Assemblies

Aqueous assembly of supramolecular polymers that use a rosette motif have been shown to require one of the edges of a pairing heterocycle be modified with a hydrophilic (e.g., charged) peripheral group to block lateral association (e.g., sheet formation) and promote water solubility. As discussed previously, both TAP and Cy can be modified with an exocyclic charged group and coassemble with the other unmodified heterocycle in water. In this study, the assembly of alternative bases into TAP-CyCo6 rosette assemblies (see Chapter 3) are initially explored (Figure 4.2) [42]. ^{1}H NMR spectroscopy is a powerful method for following monomer incorporation into supramolecular assemblies that form through a cooperative assembly mechanism [45, 46]. For these systems, the resonance signals of monomers incorporated into supramolecular polymers are invisible (i.e., broadened to baseline) as a result of slow tumbling in solution. For example, when the concentrations of TAP and CyCo6 are below 15 mM, the minimum assembly concentration (MAC) at pH 7 and 20 °C, the integrated intensity of the CyCo6 signal corresponds to the actual concentration of CyCo6. However, increasing monomer concentrations above 15 mM does not lead to an increase in CyCo6 resonance signal intensity because all additional monomers are incorporated into large assemblies (see Figure 3.11 in the previous chapter) [42].

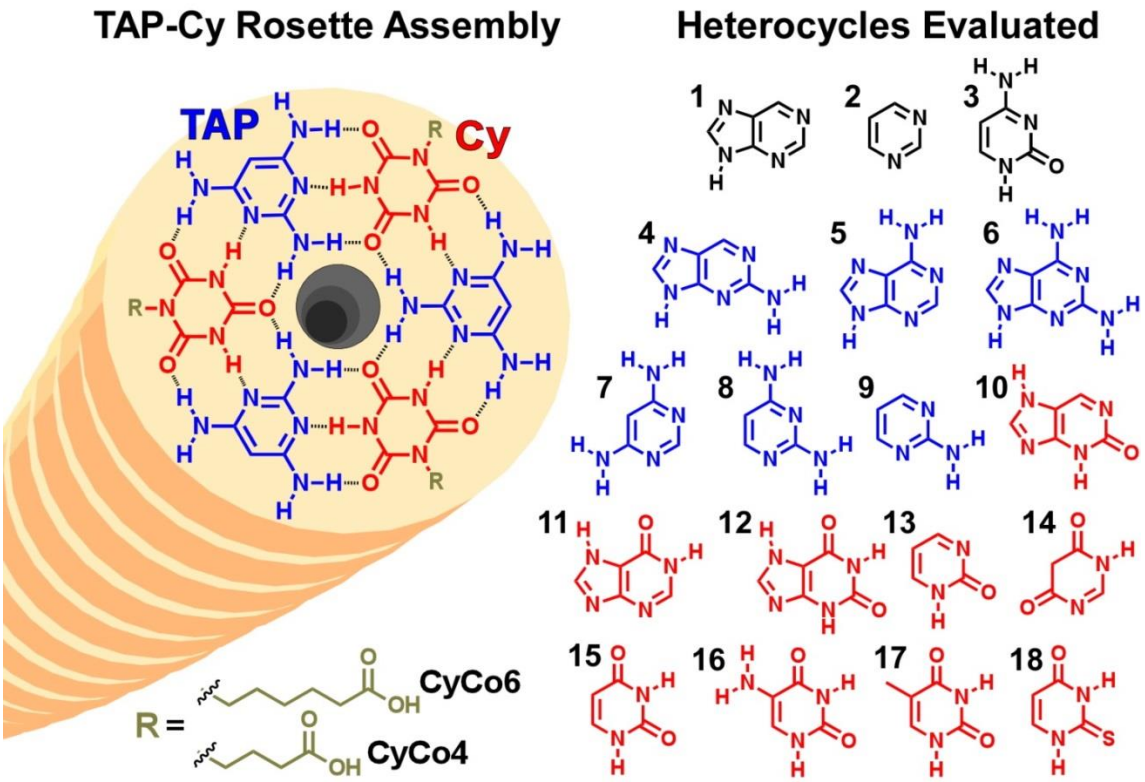


Figure 4.2 Graphical representation of the TAP-Cy rosette assembly and stacking, CyCo6 monomer and heterocycles tested for possible substitution into the rosette. **1**, purine; **2**, pyrimidine; **3**, cytosine; **4**, 2-aminopyrimidine; **5**, adenine; **6**, 2,6-diaminopyrimidine; **7**, 4,6-diaminopyrimidine; 8, 2,4,-diaminopyrimidine; **9**, 2-aminopyrimidine; **10**, 2-hydroxypurine; **11**, hypoxanthine; **12**, xanthine; **13**, 2-pyrimidinone; **14**, 2,4-dihydroxypyrimidine; **15**, uracil; **16**, 5-aminouracil; **17**, thymine; **18**, 2-thiouracil.

Screening for the ability to be incorporated into the TAP-CyCo6 assemblies was performed on 18 heterocycles possessing either the purine or pyrimidine rings system. Based on the similarity in the positioning of H-bond donor and acceptor groups, the evaluated heterocycles are categorized in Figure 4.2 as TAP-like (red), Cy-like (blue) and undefined (black). ^{1}H NMR was used to investigate the possible incorporation of alternative heterocycles into TAP-CyCo6 assemblies. In these experiments, 5 mM of each heterocycle (**1**-**18** in Figure 4.2) was added to a solution that was 30 mM in both CyCo6 and TAP. The concentration of *free*, that is unassembled, heterocycle was determined by integrating the resonance peak of a specific non-exchangeable proton. The NMR spectra used to evaluate the possible incorporation of purine ($C_5N_5H_5$, **1**) are shown in Figure 4.3. The spectrum of purine in the presence of TAP-CyCo6 assemblies is unchanged from that of the free molecule, both with respect to resonance intensity and resonance line width. Furthermore, the spectrum of purine is indifferent to the amount of TAP-CyCo6 assembled in solution, as the ^{1}H spectrum does not change from 5 °C to 40 °C, even though the amount of TAP and CyCo6 assembled in solution decreases from 24 mM to 0 mM (Figure 3.12). Thus, we conclude that purine (**1**) is not appreciably incorporated into TAP-CyCo6 assemblies.

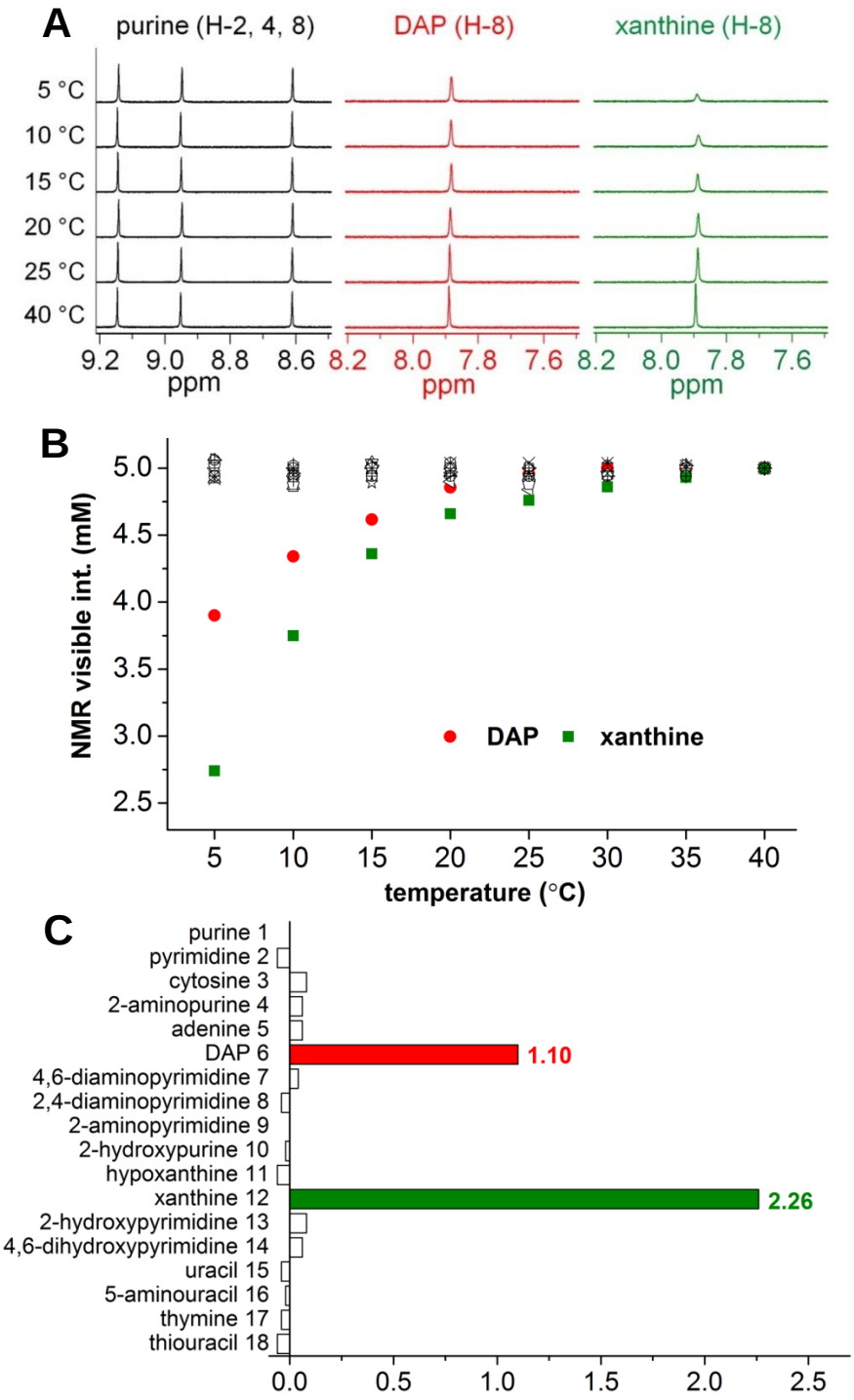


Figure 4.3 NMR analysis of solutions containing TAP and CyCo6 at 30 mM each, with competing heterocycle at 5 mM. (A) Representative ^{1}H NMR spectra of select nonexchangeable protons of purine (**1**), DAP (**6**) and xanthine (**12**) at various temperatures. (B) Plots of apparent solution-phase concentrations of each heterocycle from 5 °C – 40 °C. (C) The loss of NMR signal of each heterocycle at 5 °C, which corresponds to incorporation into TAP-CyCo6 assemblies.

Among the 18 heterocycles investigated, only DAP and xanthine (**12**, shown in Figure 3.2 as the 7H-dioxo tautomer, the favored tautomeric form) [47, 48] were identified with the ability to incorporate into TAP-CyCo6. The ^{1}H NMR spectra of these molecules exhibits reduced integrated resonance intensity in the presence of TAP-CyCo6 assemblies (Figure 4.3). Specifically, the integrated intensity of free DAP at 5 °C is 3.9 mM, indicating 1.1 mM incorporation of this molecule into the 24 mM of assembled TAP-CyCo6. The ^{1}H spectrum of free xanthine at 5 °C is 2.75 mM, indicating the incorporation of 2.25 mM of xanthine into the 24 mM of assembled TAP-CyCo6. Although DAP and xanthine show different propensities to incorporate into the TAP-CyCo6 assemblies at 5 °C, the reduction of both DAP and xanthine incorporation with increased temperature, as indicated by increased signal intensity (Figure 4.3B), is proportional to the concentration of the TAP-CyCo6 assemblies that are present at temperatures from 5 °C to 40 °C.

To provide further evidence that DAP and xanthine are entering the TAP-CyCo6 assemblies, as opposed to forming a self-structure, the two heterocycles were investigated separately by ^{1}H-NMR for incorporation as a function of increasing TAP and CyCo6 concentration was assessed quantitatively at 5 °C by NMR spectroscopy. The integrated intensity of DAP and xanthine ^{1}H resonances were found to decrease linearly with increasing TAP and CyCo6, indicating that supramolecular assemblies promote incorporation of these purines (Figure 4.4).

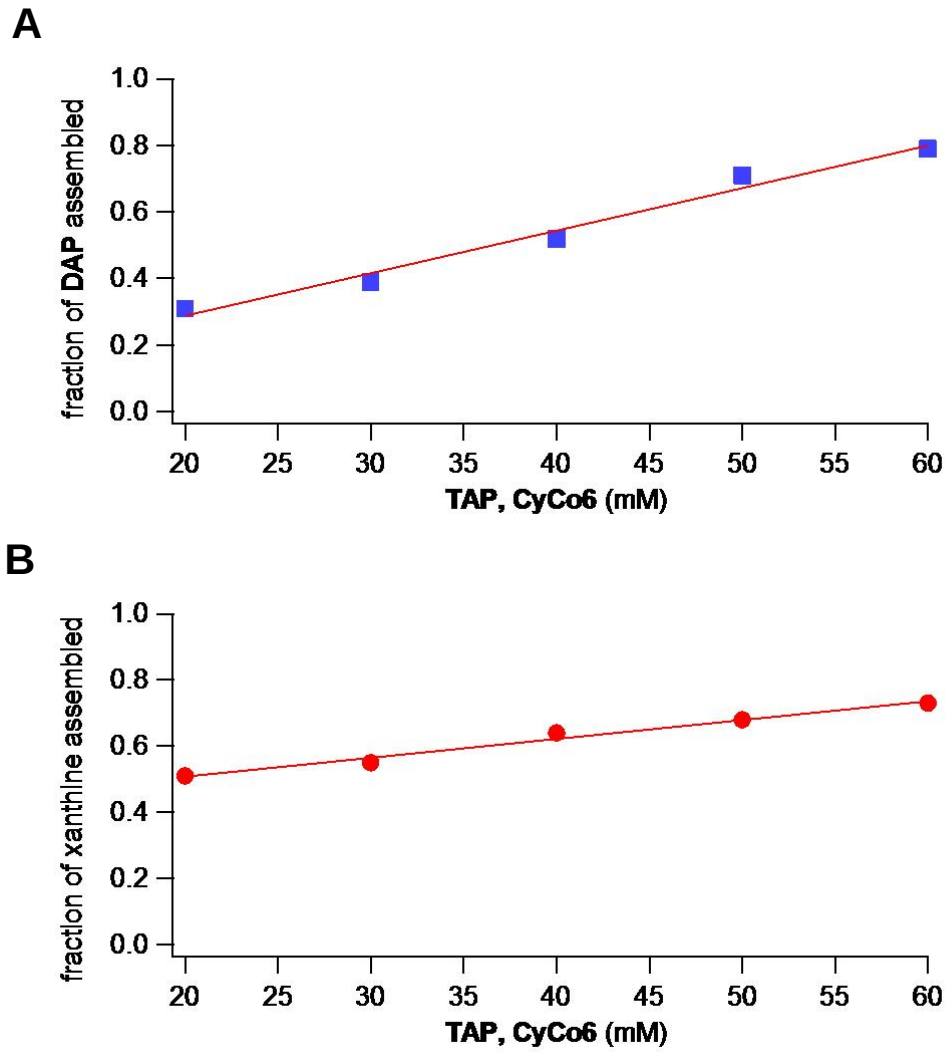


Figure 4.4 Plots showing that increasing concentration of TAP and CyCo6 promotes DAP and xanthine assembly. Plots were derived from NMR analysis of the fraction of total (A) DAP or (B) xanthine assembled as a function of increasing concentrations of TAP and CyCo6 at 5 °C. The concentration of DAP and xanthine were held constant at 5 mM for each experiment. Trend lines are shown.

There are several possible reason for the observed preferential incorporation of xanthine over DAP into the same TAP-CyCo6 assembly. This enhanced incorporation may be due to the smaller pK_a differences between H-bond donor and acceptor groups of TAP and xanthine relative to those of CyCo6 and DAP, which is known to lead to stronger H-bonding [49-52]. Additionally, the peripheral charge is decreased upon incorporation of xanthine but not DAP, which may lead to a more favorable assembly [53, 54]. However, the stacking interactions of DAP and xanthine, when substituted into the assemblies, may also contribute to the observed preferential incorporation, and predicting the influence of exocyclic groups on the energetics of stacked heterocycles in aqueous solvent is still extremely difficult even for well-defined systems, such as DNA and RNA [55]. Finally, the interactions of DAP and xanthine with themselves and with solvent molecules may also contribute to the more favorable incorporation of xanthine.

4.3.2 CyCo6 Assembles with 2,4-Diaminopyrimidine

Our combined results reveal that under the conditions tested for the screening study (i.e., 5 mM in competing heterocycle and 30 mM TAP and CyCo6) supramolecular polymers formed from hexads with TAP and Cy as recognition units will only allow substitution by heterocycles that possess the same hydrogen bonding patterns on two faces that match those of TAP or Cy. This highly restrictive condition is only satisfied by DAP and xanthine for the 18 heterocycles shown in Figure 4.1. Those monomers that exhibit full hydrogen bonding complementarity on one face but lack even a single hydrogen bond donor or acceptor on another (e.g., 2,4-diaminopyrimidine (**8**), uracil (**15**), 2-aminopurine (**4**) and 2-hydroxypurine (**10**)) would be expected to have been the next

most likely, after DAP and Xanthine, to assemble, however no incorporated into the assembly was observed under the conditions examined. To evaluate the limits of the assembly process 2,4-diaminopyrimidine was evaluated for its ability to coassemble with CyCo6 at higher concentrations.

While no gelation was observed with 2,4-diaminopyrimidine and CyCo6 at 30 mM, a solution at 200 mM in both molecules formed a weak translucent hydrogel when incubated on ice (Figure 4.5A). Attempts to make a more concentrated solution of 500 mM in both molecules failed due to poor solubility at such high concentrations. ^{1}H NMR revealed a 1:1 coassembly in a solution containing 200 mM of both 2,4-diaminopyrimidine and CyCo6 and 500 mM sodium phosphate (Figure 4.5). The assemblies were found to be highly sensitive to temperature, with the MAC changing from 80 mM at 5 °C to ca. 120 mM at 20 °C. The MAC was found to be approximately one order of magnitude greater at temperatures between 5 and 35 °C, for the 2,4-diaminopyrimidine-CyCo6 assembly compared to the TAP-CyCo6 assembly (for example, at 5 °C TAP-CyCo6 MAC is 7 mM at at 20 °C the MAC increases to 15 mM). This demonstrates the importance in maintaining 18 hydrogen bonds for assembly to occur. Additionally, taking into account the similarity of TAP to 2,6,-diaminopyrimidine, it is likely the other heterocycles evaluated in the screening that showed no incorporation into the TAP-CyCo6 assembly will not be compatible with monomeric assembly at low millimolar concentrations.

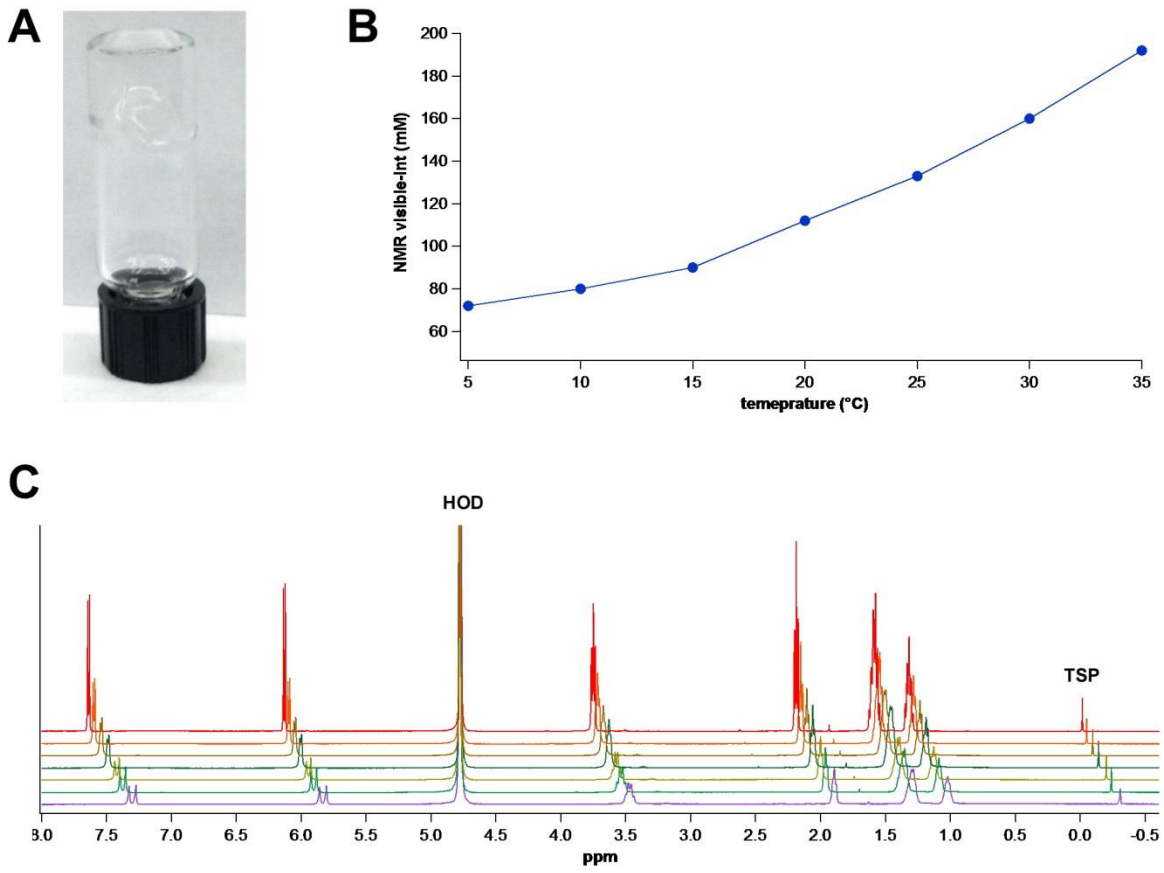


Figure 4.5 Gel inversion test and NMR analysis of 2,4-diaminopyrimidine and CyCo6 assembly. (A) Inverted bottle test showing hydrogel formation of a solution that was 200 mM in both molecules. (B) Plot of the visible amount of 2,4-diaminopyrimidine (or CyCo6) visible by NMR at various temperatures. (C) NMR spectra used to generate plot shown in (B). TSP indicates internal standard that does not interact with assemblies.

4.3.3 CyCo4 and 2,6-Diamiopurine Form Rosette-Based Supramolecular Polymers

Since screening showed incorporation of DAP into the CyCo6-TAP supramolecular polymers, we next investigated if DAP could fully substitute for TAP to form a Cy-DAP assembly motif, which would represent the first hexameric rosette assembly motif that employs both a monocyclic and bicyclic ring system. Upon combining equimoler portions of CyCo6 and DAP, a precipitate formed immediately that consisted of a 1:1 stoichiometry of each monomer (Figure 4.6 and 4.7). Modest changes in the peripheral groups of rosette-forming monomers have been shown to dramatically affect the solubility of supramolecular assemblies [19, 42, 56-58]. These observations motivated us to investigate if a Cy tagged with butynoic acid, CyCo4, could substitute for CyCo6 and enable the formation of soluble assemblies. When DAP and CyCo4 were combined a shear-thinning hydrogel appeared (Figure 4.6), which is indicative of the formation of noncovalently cross-linked supramolecular polymers. The hydrogel phase was found to be metastable with the lifetime increasing with increasing CyCo4:DAP ratio. For example, a solution that is 30 mM in both CyCo4 and DAP precipitates with 20 minutes while a solution containing 30 mM DAP and 45 mM CyCo4 persists for up to 8 hours on ice before precipitating, but could be reformed by heating to dissolve the precipitate. ^{1}H NMR analysis (Figure 4.7) of a solution containing 20 mM each of DAP and of CyCo4 at 5 °C confirmed that the stoichiometry of the soluble assemblies is 1:1.

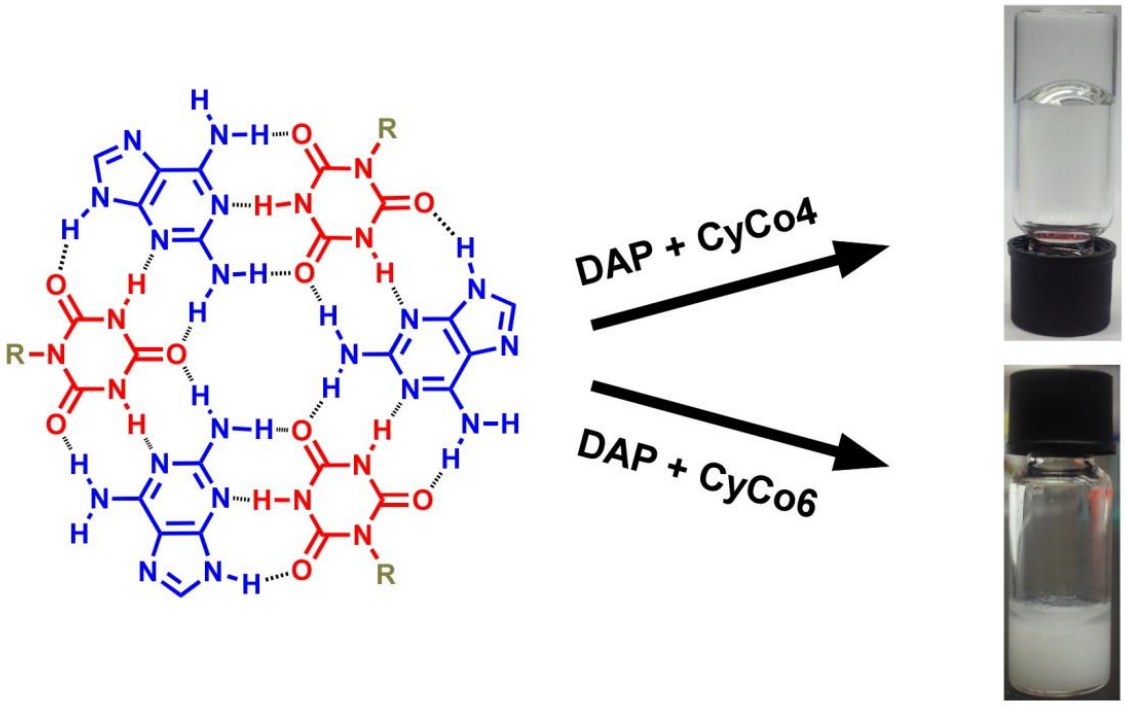


Figure 4.6 Proposed rosette assembly of DAP and Cy-tagged monomers (R groups are the same as in Figure 4.2), and inverted-bottle test of solutions containing DAP and CyCo4 or DAP and CyCo6 (60 mM of each monomer).

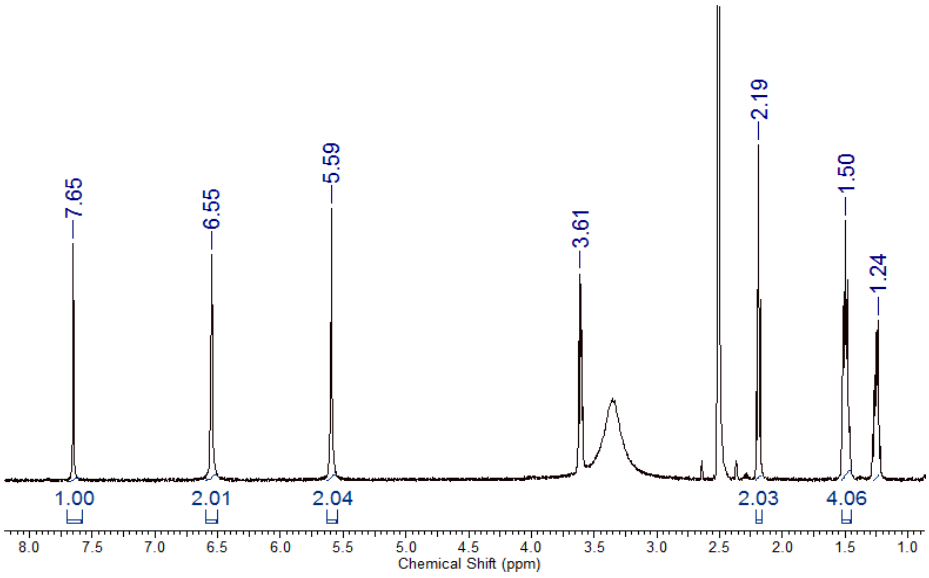


Figure 4.7 ^{1}H NMR spectra (DMSO-d_6) of DAP and CyCo6 that coprecipitated from water. The peaks at 7.65, 6.55, and 5.59 ppm are assigned to H-8, 2-NH_2, and 6-NH_2 in DAP, respectively. The peaks at 3.61, 2.19, 1.50, and 1.24 ppm are assigned to the methylene protons of CyCo6. The integrated intensity demonstrates a 1:1 stoichiometry of DAP and CyCo6 in the precipitate.

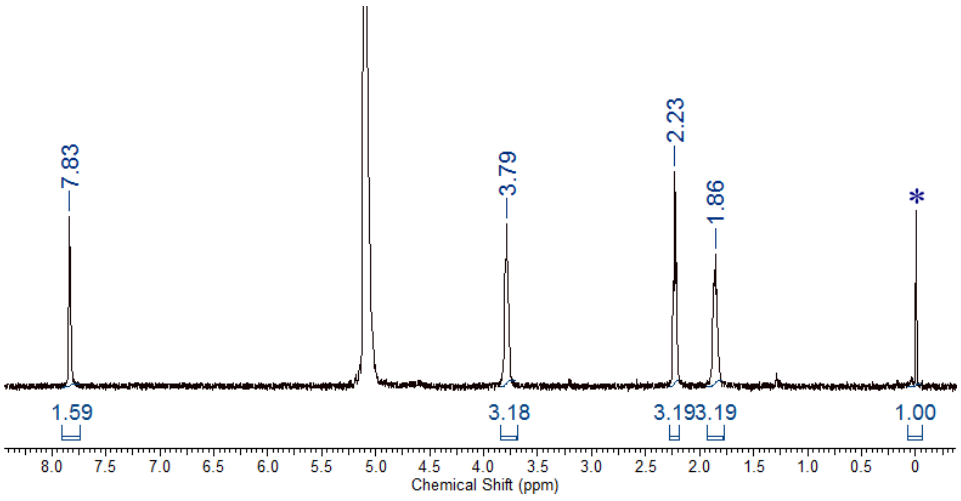


Figure 4.8 ^{1}H NMR spectra of a solution containing a 1:1 mixture of DAP and CyCo4 (20 mM each) at 5 °C. The peak at 7.83 ppm is assigned to H-8 of DAP, which shows 7.95 mM DAP free in the solution (12.05 mM DAP assembled). The peaks at 3.79, 2.23, and 1.86 ppm are assigned to the methylene protons of CyCo4. The integrated intensity presents 7.95 mM CyCo4 free in the bimolecular system (12.05 mM CyCo4 assembled). TSP is labeled (*).

UV-vis absorption spectroscopy is a convenient method to monitor interactions between the monomers of these non-covalent polymers. The UV spectrum of **DAP** in the gel phase with **CyCo4** is red-shifted relative to the spectrum of free **DAP** (Figure 4.9A). This spectral change is indicative of intermolecular π-π electron interactions that characterize J-type associations typical of planar molecules [59]. The spectroscopic behavior of 1:1 mixtures of DAP-CyCo4 suggests that intermolecular stacking interactions occur at concentrations above 25 mM (Figure 4.9B). Analysis of the UV-vis and NMR data reveal that the minimal assembly concentration for DAP-CyCo4 is 25 mM at 20 °C and 8 mM at 5 °C. For comparison, and as stated above, the TAP-CyCo6 system has a minimal assembly concentration of 15 mM at 20 °C and 6 mM at 5 °C.

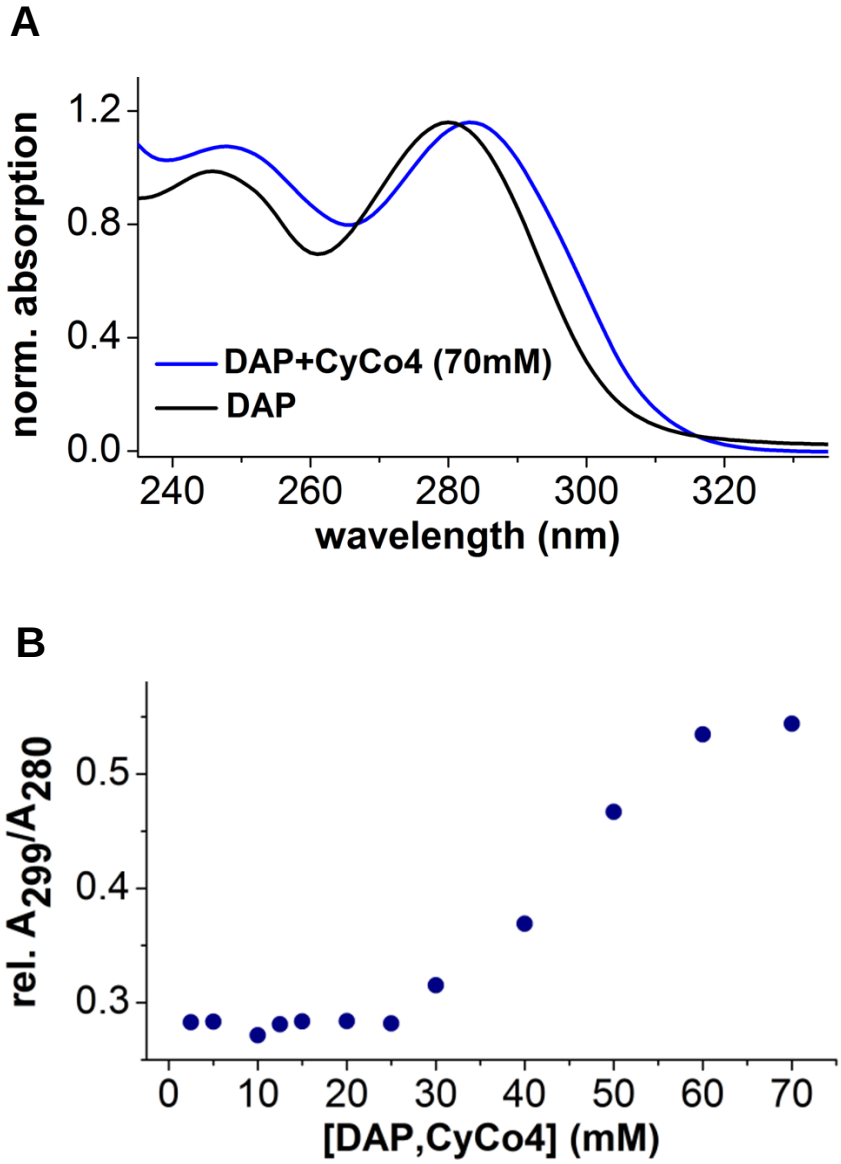


Figure 4.9 UV-vis analysis of solutions containing DAP and CyCo4 and various concentrations. (A) UV spectra of solutions with only DAP (black curve) and 1:1 DAP/CyCo4 mixtures at 70 mM each (blue curve). (B) Plot of the absorption ratio of $A_{299\ nm}$/ $A_{280\ nm}$ as a function of monomer concentration in a 1:1 mixture of DAP and CyCo4. All of the spectra were taken at 20 °C. CyCo4 does not absorb light within the UV range displayed.

DAP-CyCo4 assemblies were visualized by atomic force microscopy (AFM), which revealed bundled supramolecular polymers with heights that ranged on average between 2 to 5 nm (Figure 4.10). These observations are fully consistent with DAP and CyCo4 assembling into rosettes that stack to form linear supramolecular polymers, which is in agreement with the presence of fibers with widths of approx. 2 nm (Figure 4.10). The formation of high-order, bifurcating bundles are also consistent with the existence of a supramolecular polymer network that composes the gel matrix. Optical polarized microscopy (POM) was used to further evaluate the hierarchical ordering of the assemblies, showing bifringence that is indicative of a lyotropic nematic liquid crystal (Figure 4.10). Similar features were also observed for gelled solutions containing CyCo6 and TAP. While other hexameric rosette based systems have been reported to organize into a lyotropic liquid crystalline phases [25, 60, 61], the DAP-Cy motif is the first reported to generate a liquid crystal in water, a finding that may prove useful in materials development [62].

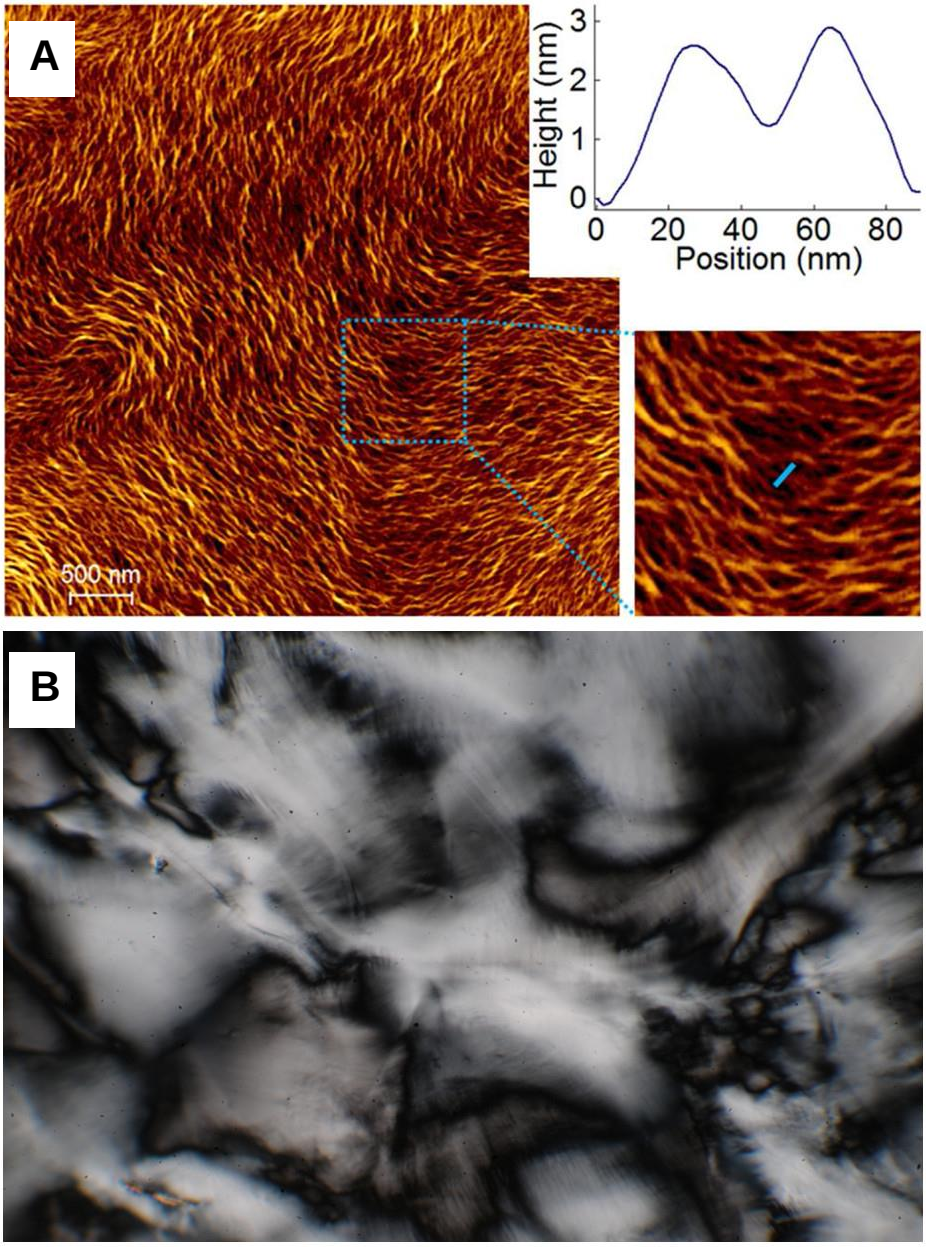


Figure 4.10 AFM and POM analysis of DAP-CyCo4. (A) AFM topographic image of self-assembled DAP-CyCo4 fibers. Insert shows the height profiles of single fibers delineated by the blue lines in the zoom panel. Sample was 20 mM in each molecule and cooled on ice prior to application to the mica surface. (B) Nematic texture of a DAP-CyCo4 solution observed by polarized microscopy using cross-polarizers. Sample was 30 mM in DAP and 45 mM in CyCo6 and analyzed at room temperature.

It was initially anticipated that the heterocycles with purine ring systems would assemble better than the pyrimidines as the increased rosette surface area was expected to enhance assembly, thus lowering the MAC, as a result of favorable stacking interactions [7, 40]. If this was the case DAP and xanthine would have a greater ratio of monomer assembled than TAP and CyCo6, respectively. However, from the initial screening experiment at 5 °C the ratio of total DAP assembled is 0.22 and for xanthine it is 0.45, while the ratio of total TAP and CyCo6 assembled is 0.8 (Note that if the purines had the same assembly potential as TAP or CyCo6 the ratio of total monomer assembled would be the same). When the bimolecular assembly of DAP and CyCo4 was evaluated at 5 °C the MAC was found to be 8 mM corresponding to an increase in the ratio of DAP assembled to 0.6. The poorer assembly ability of both DAP and xanthine compared to TAP and CyCo6 in the trimolecular assembly, and the DAP-CyCo4 bimolecular assembly compared to the assembly between TAP and CyCo6 may be a result of non-ideal surface matching. It is known, particularly with regards to intercalators for nucleic acid duplexes, triplexes and quadruplexes, that the stacking of aromatic surfaces in water is more favorable when the surfaces are of a similar size [63]. In the trimolecular assemblies, the rosettes that incorporate a purine would have poorer overlap with neighboring TAP-CyCo6 rosettes, which could provide an unfavorable penalty for assembly. Overlap would be enhanced in the bimolecular assembly of a purine and triazine and lead to an increase in the monomer assembled as seen with DAP-CyCo4 as compared to DAP-TAP-CyCo6. Therefore, this proposal is consistent with the increase in assembled DAP going from the trimolecular to the biomolecular assembly, and the

observation that TAP and CyCo6 have the greatest affinity for one another since surface overlap between rosettes formed from these molecules is maximized.

4.4 Conclusion

In summary, we have evaluated eighteen purine and pyrimidine heterocycles for their ability to be incorporated into a supramolecular assembly formed by TAP-Cy hexameric rosettes. Under the conditions tested, the incorporation of competing monomers is found to be restricted to monomers with H-bond donor/acceptor groups that fully match those of TAP and Cy. At sufficiently higher concentrations 2,4-diaminopyrimidine was found to assemble with CyCo6, demonstrating that loss of a hydrogen bond, and corresponding stacking interactions, can lead to a significant loss in the ability of a monomer to assemble. Additionally, our observation supports the benefit in maintaining full overlap between stacked hydrophobic rosette surfaces to enhance molecular assembly in water. The experimental results presented also provided strong evidence for the ability of DAP and xanthine to co-assemble with the TAP-CyCo6 system to form trimolecular assemblies. Building on these observations, it was discovered that DAP can spontaneously assemble with CyCo4 in equimolar ratio to form a hydrogel with liquid crystalline properties, which is the first demonstration of a purine-based rosette assembly system. In contrast, DAP forms soluble assemblies with CyCo6 only in the presence of TAP, an observation that again demonstrates that monomer heterogeneity within rosette assemblies, this time with the recognition elements, can enhance the stability of soluble supramolecular polymers [42].

The observed selectivity of hexad assemblies provides a potential resolution to the question of how the original nucleobases of nucleic acids, if different from the extant bases [41, 64, 65], could have abiotically organized themselves and simultaneously excluded other similar heterocycles, prior to polymerization. As discussed in Chapter 1, the chemical space for the number of heterocycles that could have competed with the original proto-RNA heterocycles is great. Heterocycles capable of their own mutual selection would have had an advantage for incorporation into an RNA-like polymer within likely complex prebiotic environments. Therefore, under the hypothesis that the first nucleobases organized themselves in solution prior to polymerization, we see from the results presented in this chapter that there are likely only a few candidate proto-nucleobases that fit this criterion.

4.5 References

1. Aida, T., E.W. Meijer, and S.I. Stupp, *Functional supramolecular polymers.* Science, 2012. **335**: p. 813-817.
2. Dong, R., et al., *Supramolecular hydrogels: synthesis, properties and their biomedical applications.* Biomaterials Science, 2015: p. DOI: 10.1039/C4BM00448E.
3. Beingessner, R.L., et al., *Bioactive Rosette Nanotubes for Bone Tissue Engineering and Drug Delivery.* 2013: p. 313-357.
4. Hirst, A.R., et al., *High-tech applications of self-assembling supramolecular nanostructured gel-phase materials: from regenerative medicine to electronic devices.* Angew Chem Int Ed Engl, 2008. **47**: p. 8002-8018.
5. Hartgerink, J.D., E. Beniash, and S.I. Stupp, *Self-assembly and mineralization of peptide-amphiphile nanofibers.* Science, 2001. **294**: p. 1684-1688.
6. Zurcher, D.M. and A.J. McNeil, *Tools for Identifying Gelator Scaffolds and Solvents.* J Org Chem, 2015.

7. De Greef, T.F., et al., *Supramolecular polymerization.* Chem Rev, 2009. **109**: p. 5687-5754.

8. Wojtecki, R.J., M.A. Meador, and S.J. Rowan, *Using the dynamic bond to access macroscopically responsive structurally dynamic polymers.* Nature Mater., 2011. **10**: p. 14-27.

9. Rehm, T.H. and C. Schmuck, *Ion-pair induced self-assembly in aqueous solvents.* Chem Soc Rev, 2010. **39**: p. 3597-611.

10. Banerjee, S. and C. Schmuck, *Formation of Hydrogen-Bonded Self-assembled Structures in Polar Solvents.* 2015. **87**: p. 187-225.

11. Leenders, C.M.A., et al., *From supramolecular polymers to hydrogel materials.* Mater Hori., 2014. **1**: p. 116.

12. Frisch, H. and P. Besenius, *pH-Switchable Self-Assembled Materials.* Macromol Rapid Commun, 2014.

13. Buerkle, L.E. and S.J. Rowan, *Supramolecular gels formed from multi-component low molecular weight species.* Chem Soc Rev, 2012. **41**: p. 6089-6102.

14. Frisch, H., et al., *pH-Regulated Selectivity in Supramolecular Polymerizations: Switching between Co- and Homopolymers.* Chemistry, 2015. **21**: p. 3304-3309.

15. Fathalla, M., et al., *Base-pairing mediated non-covalent polymers.* Chem Soc Rev, 2009. **38**: p. 1608-1620.

16. Cheng, S., et al., *Nucleobase Self-Assembly in Supramolecular Adhesives.* Macromolecules, 2012. **45**: p. 805-812.

17. Largy, E., et al., *A pyrimidopyrimidine Janus-AT nucleoside with improved base-pairing properties to both A and T within a DNA duplex: the stabilizing effect of a second endocyclic ring nitrogen.* Chemistry, 2014. **20**: p. 1495-1499.

18. Rivera, J.M., *Functional Assemblies made from Supramolecular G-quadruplexes.* 2012: p. 15-27.

19. Durmus, A., et al., *Synthesis of N-substituted Pyrido[4,3-d]pyrimidines for the Large-Scale Production of Self-Assembled Rosettes and Nanotubes.* J Org Chem, 2013. **78**: p. 11421-11426.

20. Mascal, M., et al., *Programming a Hydrogen-Bonding Code for the Specific Generation of a Supermacrocycle.* Angew Chem Int Ed, 1996. **35**: p. 2204-2206.

21. Asadi, A., B.O. Patrick, and D.M. Perrin, *Janus-AT bases: synthesis, self-assembly, and solid state structures.* J Org Chem, 2007. **72**: p. 466-475.

22. Zhao, H., et al., *Different superstructures formed by Janus-type nucleosides.* Chem Comm, 2012. **48**: p. 6097-6099.

23. Borzsonyi, G., et al., *Rosette nanotubes with 1.4 nm inner diameter from a tricyclic variant of the Lehn-Mascal G--C base.* Chem Comm, 2010. **46**: p. 6527-6529.

24. Kolotuchin, S.V. and S.C. Zimmerman, *Self-Assembly Mediated by the Donor−Donor−Acceptor·Acceptor−Acceptor−Donor (DDA·AAD) Hydrogen-Bonding Motif: Formation of a Robust Hexameric Aggregate.* J Am Chem Soc, 1998. **120**: p. 9092-9093.

25. Yagai, S., et al., *Self-Organization of Hydrogen-Bonding Naphthalene Chromophores into J-type Nanorings and H-type Nanorods: Impact of Regioisomerism.* Angew Chem Int Ed, 2012. **124**: p. 6747-6751.

26. Seto, C.T. and G.M. Whitesides, *Self-assembly based on the cyanuric acid melamine lattice.* J Am Chem Soc, 1990. **112**: p. 6409-6411.

27. Whitesides, G.M., et al., *Noncovalent Synthesis: Using Physical-Organic Chemistry To Make Aggregates* Acc Chem Res, 1994. **28**: p. 37-44.

28. Choi, I.S., et al., *Self-assembly of hydrogen-bonded polymeric rods based on the cyanuric acid center dot melamine lattice.* Chem Mat, 1999. **11**: p. 684-690.

29. Lehn, J.-M., et al., *Molecular recognition directed self-assembly of ordered supramolecular strands by cocrystallization of complementary molecular components.* J Chem Soc, Chem Commun, 1990: p. 479.

30. Prins, L.J., et al., *An enantiomerically pure hydrogen-bonded assembly.* Nature, 2000. **408**: p. 181-184.

31. Yagai, S., T. Karatsu, and A. Kitamura, *Photoresponsive melamine-barbiturate hydrogen-bonded assembly.* Chem Commun, 2003: p. 1844.

32. Yagai, S., et al., *Toroidal Nanoobjects from Rosette Assemblies of Melamine-Linked Oligo(p-phenyleneethynylene)s and Cyanurates.* Angew Chem Int Ed, 2008. **120**: p. 4769-4772.

33. Rakotondradany, F., et al., *Hydrogen-bond self-assembly of DNA-analogues into hexameric rosettes.* Chem Comm, 2005: p. 5441-5443.

34. Fenniri, H., et al., *Helical rosette nanotubes: Design, self-assembly, and characterization.* J Am Chem Soc, 2001. **123**: p. 3854-3855.

35. Moralez, J.G., et al., *Helical rosette nanotubes with tunable stability and hierarchy.* J Am Chem Soc, 2005. **127**: p. 8307-8309.

36. Bozsonyi, G., et al., *Water-soluble J-type rosette nanotubes with giant molar ellipticity.* J Am Chem Soc, 2010. **132**: p. 15135-15139.

37. Fenniri, H., et al., *Entropically driven self-assembly of multichannel rosette nanotubes.* Proc Natl Acad Sci U S A, 2002. **99**: p. 6487-6492.

38. Zhang, L., et al., *Arginine-glycine-aspartic acid modified rosette nanotube-hydrogel composites for bone tissue engineering.* Biomaterials, 2009. **30**: p. 1309-1320.

39. Borzsonyi, G., et al., *Water-soluble J-type rosette nanotubes with giant molar ellipticity.* J Am Chem Soc, 2010. **132**: p. 15136-15139.

40. Cafferty, B.J., et al., *Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues.* J Am Chem Soc, 2013. **135**: p. 2447-2450.

41. Chen, M.C., et al., *Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle.* J Am Chem Soc, 2014. **136**: p. 5640-5646.

42. Cafferty, B.J., et al., *Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK_a matching of biomimetic monomers.* Chem Sci, 2014. **5**: p. 4681-4686.

43. Hager, K., A. Franz, and A. Hirsch, *Self-assembly of chiral depsipeptide dendrimers.* Chem Eur J, 2006. **12**: p. 2663-2679.

44. Andresen, G., et al., *Regioselective addition of Grignard reagents to a 2-oxopurinium salt.* Tetrahedron, 1995. **51**: p. 3655-3664.

45. Hirst, A.R., et al., *Low-molecular-weight gelators: elucidating the principles of gelation based on gelator solubility and a cooperative self-assembly model.* J Am Chem Soc, 2008. **130**: p. 9113-9121.

46. Duncan, D.C. and D.G. Whitten, *^{1}H NMR Investigation of the Composition, Structure, and Dynamics of Cholesterol−Stilbene Tethered Dyad Organogels.* Langmuir, 2000. **16**: p. 6445-6452.

47. šponer, J. and J. Leszczyński, *Tautomerism of xanthine: The second-order MØller-Plesset study.* Struct Chem, 1995. **6**: p. 281-286.

48. Callahan, M.P., et al., *IR-UV double resonance spectroscopy of xanthine.* Phys Chem Chem Phys, 2007. **9**: p. 4587-91.

49. Perrin, C.L. and J.B. Nielson, *"Strong" hydrogen bonds in chemistry and biology.* Annu Rev Phys Chem, 1997. **48**: p. 511-544.

50. Chen, J., et al., *Short, Strong Hydrogen Bonds in the Gas Phase and in Solution: Theoretical Exploration of pKaMatching and Environmental Effects on the Strengths of Hydrogen Bonds and Their Potential Roles in Enzymatic Catalysis.* J Org Chem, 1998. **63**: p. 4611-4619.

51. Acharya, P., et al., *Measurement of nucleobase pK_a values in model mononucleotides shows RNA-RNA duplexes to be more stable than DNA-DNA duplexes.* J Am Chem Soc, 2004. **126**: p. 2862-2869.

52. Engelhart, A.E., T.H. Morton, and N.V. Hud, *Evidence of strong hydrogen bonding by 8-amino-guanine.* Chem Comm, 2009: p. 647-649.

53. Schaefer, C., et al., *Controlling the Cooperativity in the Supramolecular Polymerization of Ionic Discotic Amphiphiles via Electrostatic Screening.* ACS Macro Letters, 2012. **1**: p. 830-833.

54. Yu, Y., et al., *Tunable thermoassociation of binary guanosine gels.* J Phys Chem B, 2008. **112**: p. 1130-1134.

55. Parker, T.M., et al., *Quantum-Mechanical Analysis of the Energetic Contributions to pi Stacking in Nucleic Acids versus Rise, Twist, and Slide.* J Am Chem Soc, 2013. **135**: p. 1306-1316.

56. Whitesides, G.M., et al., *Noncovalent Synthesis: Using Physical-Organic Chemistry To Make Aggregates.* Acc Chem Res, 1995. **28**: p. 37-44.

57. Marsh, A., M. Silvestri, and J.-M. Lehn, *Self-complementary hydrogen bonding heterocycles designed for the enforced self-assembly into supramolecular macrocycles.* Chem Comm, 1996(13): p. 1527.

58. Mascal, M., et al., *The G−C DNA Base Hybrid: Synthesis, Self-Organization and Structural Analysis.* J Org Chem, 1999. **64**: p. 8479-8484.

59. Jelley, E.E., *Spectral absorbance and fluorescence of dyes in their molecular state.* Nature, 1936. **138**: p. 1009-1010.

60. Yagai, S., et al., *Hierarchical organization of photoresponsive hydrogen-bonded rosettes.* J Am Chem Soc, 2005. **127**: p. 11134-11139.

61. Piermattei, A., et al., *Induction of liquid crystallinity by self-assembled molecular boxes.* Angew Chem Int Ed, 2006. **45**: p. 7543-7546.

62. Kato, T., N. Mizoshita, and K. Kishimoto, *Functional liquid-crystalline assemblies: self-organized soft materials.* Angew Chem Int Ed, 2005. **45**: p. 38-68.

63. Buckley, R., et al., *Molecular Recognition of Watson-Crick-Like Purine-Purine Base Pairs.* ChemBioChem, 2011. **12**: p. 2155-2158.

64. Hud, N.V., et al., *The origin of RNA and 'My Grandfather's Axe'.* Chem Biol, 2013. **20**: p. 466-474.

65. Cafferty, B.J. and N.V. Hud, *Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology.* Curr Opin Chem Biol, 2014. **22**: p. 146-157.

CHAPTER 5

CONTROLLING THE LENGTH OF SUPRAMOLECULAR POLYMERS USING PLANAR POLYCYCLIC MOLECULES AS NONCOVALENT TERMINATORS[e]

5.1 Introduction

The dynamic nature of supramolecular polymers, which results from the reversible, noncovalent and bidirectional interactions between the monomers, provides these systems with unique properties that set them apart from covalent polymers [1]. As described in the three preceding chapters, environmental changes (e.g., temperature and solution pH) can lead to a rapid and reversible response in the amount of monomer assembled into supramolecular polymers. While this equilibrium can be altered in a predictable fashion, supramolecular polymerization systems presently lack the high level of control over polymer length and polydispersity that is commonly achieved during the synthesis of covalent polymers [2]. The development of efficient strategies to control the degree of polymerization of small molecules into supramolecular polymers will ultimately be important to expand the utility of functional supramolecular polymers and corresponding soft materials [2-4].

Great progress has been made in the programmable self-assembly of small molecules into various architectures such as spheres, rods, cylinders, with increasing size and complexity through rational design of the molecular building blocks [5-7]. Recently,

e. Work presented in this chapter was performed in collaboration with Dr. Suneesh C. Karunakaran.

particularly within the last year, researchers have begun to develop methodologies to achieve similar control over the degree of polymerization of linear noncovalent assemblies. Strategies to achieve supramolecular polymer length control include designing monomers that frustrate growth [8], kinetically controlled polymerization [9], varying the enantiomeric excess of chiral self-assembling monomers [10], homogeneous seeding [11], or by employing tailored initiators [12]. However, strategies for governing assembly length that do not affect monomer composition or initiation have yet to be described. An alternative approach using noncovalent capping agents, which can be considered to have similar qualities to terminators in covalent polymer chemistry, could also be envisioned as a mechanism to control supramolecular polymer chain length under equilibrium conditions.

In this chapter we investigate the use of small molecules to control the degree of polymerization and the bulk physical properties of the CyCo6-TAP assembly system [13]. We propose that polycyclic planar molecules may effectively behave as noncovalent terminators if they can frustrate elongation upon stacking with the planar rosette surfaces of the supramolecular polymers. Several charged, polycyclic aromatic molecules were investigated for their ability to tune the length of these supramolecular polymers in a concentration-dependent and molecule-specific manner through our proposed end-capping mechanism. These results demonstrate a new strategy to precisely control the size of supramolecular polymers in a way that is similar to termination processes typical employed during chain growth polymerization.

5.2 Experimental Procedures

5.2.1 Materials

2,4,6-triaminopyrimidine (TAP) was purchased from Acros Organic. All polycyclic small molecules used in the study are shown in Figure 5.4 and were used as received. Synthesis of 1-(5-Carboxypentyl)-1,3,5-triazin-2,4,6-trion (**CyCo6**) is described in Chapter 3.

5.2.2 Assembly Formation

Assemblies were formed by adding CyCo6 to a solution containing an equivalent amount of TAP and 200 mM sodium phosphate buffered at pH 7; pH was adjusted with NaOH or HCl. Unless otherwise noted, assembly solutions were equilibrated and analyzed at 20 °C. CyCo6 purity were determined by NMR and LCMS. NMR analysis for molecular characterization and precipitate composition was performed in DMSO-d6 on a Varian Mercury 400 MHz NMR.

5.2.3 Spectroscopic Analysis

Spectroscopic analysis of the assemblies was performed using UV-vis and ^{1}H NMR. UV-vis analysis was carried out on an Agilent 8453 spectrophotometer equipped with an 89090A temperature controller. Cells of different path lengths (0.1 and 0.01 mm) were used depending on the concentration of the sample to maintain an optical density below 1.2. ^{1}H NMR spectra were collected on a Bruker DRX-500 500 MHz NMR and were the sum of 64 transients. All assembly solutions analyzed by NMR were dissolved

in D_2O and spectra were collected using the WATERGATE pulse sequence. TSP was used as an internal standard. Dynamic light scattering was performed on a DynaPro-99 equipped with a temperature controller.

5.2.4 Microscopy

AFM imaging was performed on freshly cleaved mica that was pre-activated with $MgCl_2$ with 1-2 h incubation. The mica substrate was rinsed with water and dried under N_2. Solutions containing the assemblies were incubated on ice just prior to deposition. A 2 µl sample of the assembly solution spread on the mica surface with N_2 gas. AFM imaging was performed with a Nanoscope IIIa (Digital Instruments) in tapping mode, using Si tips (MicroMatch, 16 N/m). Electron microscopy was performed on a JEOL100CX.

5.3 Results and Discussion

5.3. The Length of TAP-CyCo6 Assemblies Decreases in the Presence of Ethidium

The rosette surface is formed through the assembly of aromatic heterocycles that mimic the base-pairing interactions of nucleic acids. This similarity inspired us to first investigate the potential impact of the well-known nucleic acid intercalator ethidium bromide (EtBr) on TAP-CyCo6 assemblies. The mode of binding to nucleic acids by this cationic planar, polycyclic heterocycle is primarily through stacking interaction between the planar surfaces of the small molecule and those of two base pairs of a nucleic acid duplex. To test if EtBr affects the bulk properties of the hydrogel, the small molecule was added to an already gelled solution of TAP-CyCo6 at 30 mM total concentration. Using vial inversion, loss in the elastic strength of the gel was observed at concentrations of EtBr between 0.25 mM to 0.75 mM, and complete loss of gelation was observed above 1 mM (Figure 5.1A). These results are consistent with the destabilization of the three dimensional gel matrix due to supramolecular fiber shortening which would lead to a decrease in the average number of crosslinks per supramolecular polymer and thus a diminished elastic strength (see Figure 5.1B) [14].

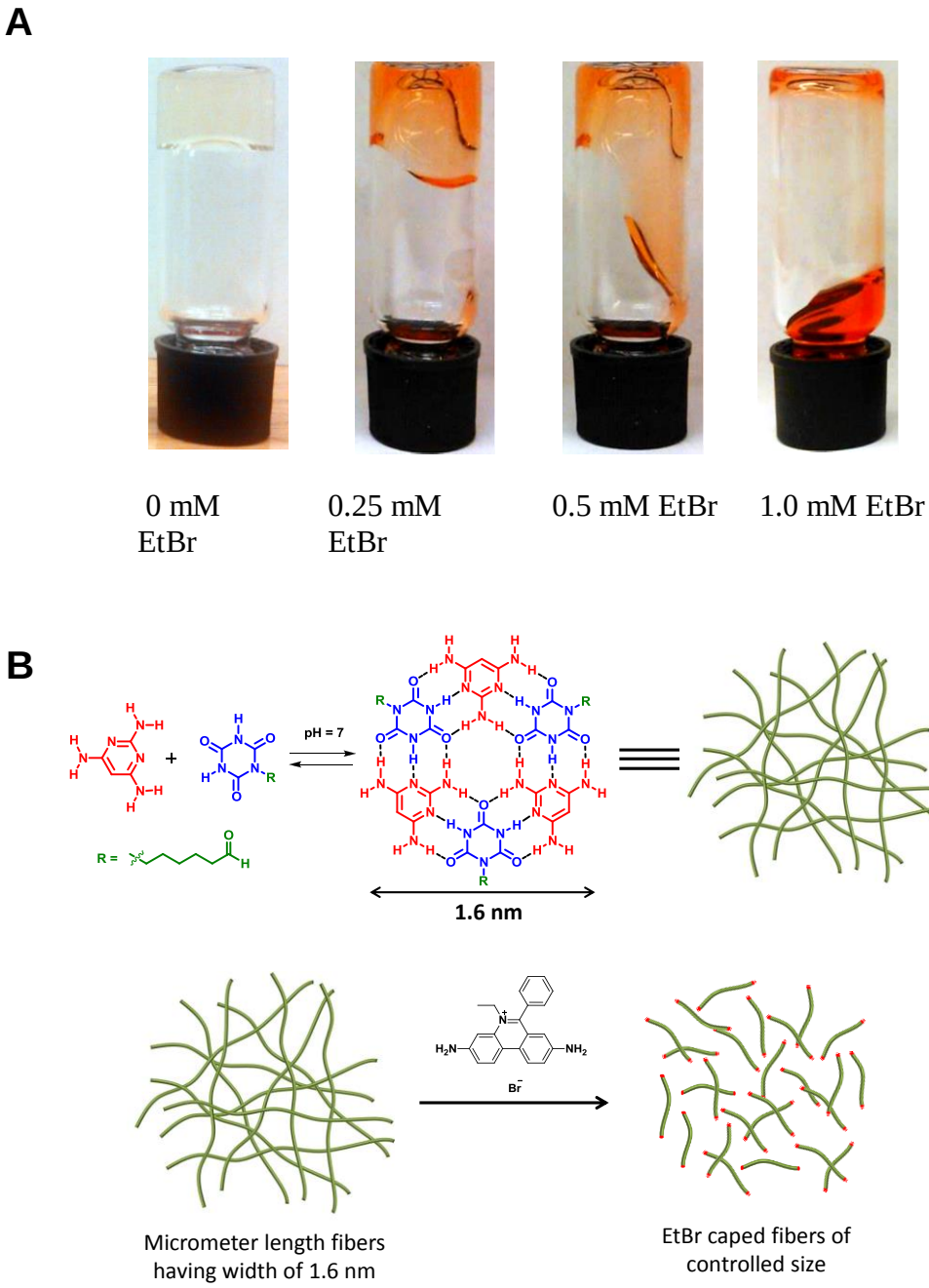


Figure 5.1 Ethidium shortens TAP-CyCo6 assemblies in a concentration dependent manner, which may be due to supramolecular polymer end-capping. (A) Vial inversion test of solutions of CyCo6 and TAP at 30 mM each and varying concentrations of EtBr. (B) Graphical representation depicting the TAP-CyCo6 rosettes, the hydrogel matrix and proposed fiber shortening by end capping of EtBr which would lead to subsequent loss of the hydrogel matrix.

AFM was next used in order to better understand the effect of EtBr on the morphology of the CyCo6-TAP supramolecular polymers. The TAP-CyCo6 supramolecular assembly alone showed micron length linear structures with heights of approximately 2 nm, consistent with the predicted width of the rosette assembly. When various concentrations of EtBr were added to solutions with the same concentration of CyCo6 and TAP, we observed a marked reduction in the length of the supramolecular polymer that was proportional to the concentration of EtBr. For example, when AFM was performed on a sample containing 30 mM TAP-CyCo6 and 1 mM EtBr, supramolecular polymers were observed with an average length of 700 nm that also possessed a very narrow polydispersity (no fibers were observed greater than 1 micron) (Figure 5.2). Similarly, when the concentration of EtBr was further increased to 2.5 mM, we observed a further reduction in the length of fibers to an average length of 360 nm. It is noted that only the length of the assemblies was changed as the diameter of the supramolecular polymers remained at 2 nm. To provide further evidence that the change in length of the supramolecular polymers resulted from the presence of EtBr and not due to sample preparation, when the same solution containing 1 mM EtBr was analyzed by TEM, 700 nm fibers were again observed (Figure 5.2D).

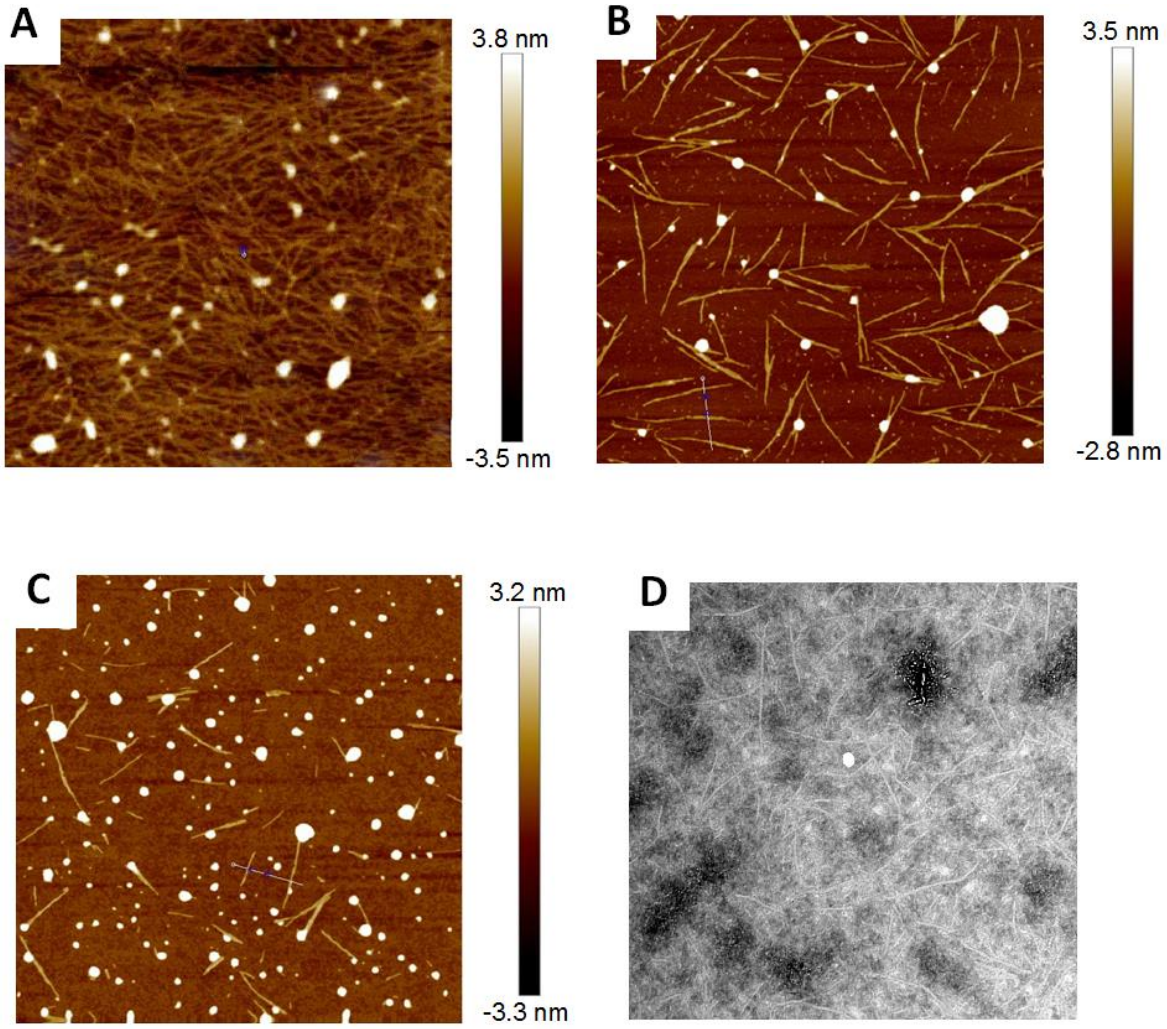


Figure 5.2 AFM and TEM images showing ethidium dependent decrease in TAP-CyCo6 assembly length. (A-C) Atomic force microscopic (AFM) images of self-assembled TAP-CyCo6 supramolecular polymers in the presence of (A) 0; (B) 1.0; (C) 2.5 mM EtBr, respectively. (D) Transmission electron microscope (TEM) image of the same solution shown in panel B (i.e., 1 mM EtBr). Each micrograph is 3 x 3 µm. All solutions originally consisted of 30 mM TAP and CyCo6. Data collected by Dr. Suneesh Karunakaran.

5.3.2 Ethidium Does Not Change the TAP-CyCo6 Minimal Assembly Concentration

To ensure that loss of gelation was due to a decrease in the length of the supramolecular polymers and not a result of EtBr inhibiting the polymerization process or amount of monomer incorporated into the assemblies, we next investigated amount of monomer assembled in the presence and absence of EtBr. Supramolecular polymers that follow the cooperative assembly mechanism are distinguished by what can be considered a phase transition at a minimal assembly concentration (MAC), below which only monomers exist in solution and above which both large assemblies and monomers (at the MAC) coexist in solution. ^{1}H NMR can effectively be used to determine the amount of monomer incorporated into supramolecular polymers that follow a cooperative mechanism [15, 16]. As reported earlier, the MAC of TAP-CyCo6 at 20 °C is 15 mM, meaning that under these conditions, assemblies will only be formed from monomer in solutions that exceeds this concentration. When the MAC was determined for TAP-CyCo6 assemblies (30 mM each) in the presence of 1-10 mM of EtBr no change in the MAC was observed. These experiments confirm that under these conditions EtBr is not affecting the assembly of CyCo6 and TAP in solution.

Additionally, at 1 mM EtBr the width and chemical shift of the EtBr resonances do not change in the presence of assembled TAP-CyCo6 (Figure 5.3B). This observation indicates that the association of EtBr with the assemblies is weaker than the interaction between TAP and CyCo6, which is consistent with the stoichiometry that should be expected for binding to supramolecular polymers. That is, if 6 EtBr (1 mM total) and 2000 rosettes (5 mM total *assembled*) are present per supramolecular polymer, 0.015 mM EtBr would be assembled, which is below the detection limit of the NMR.

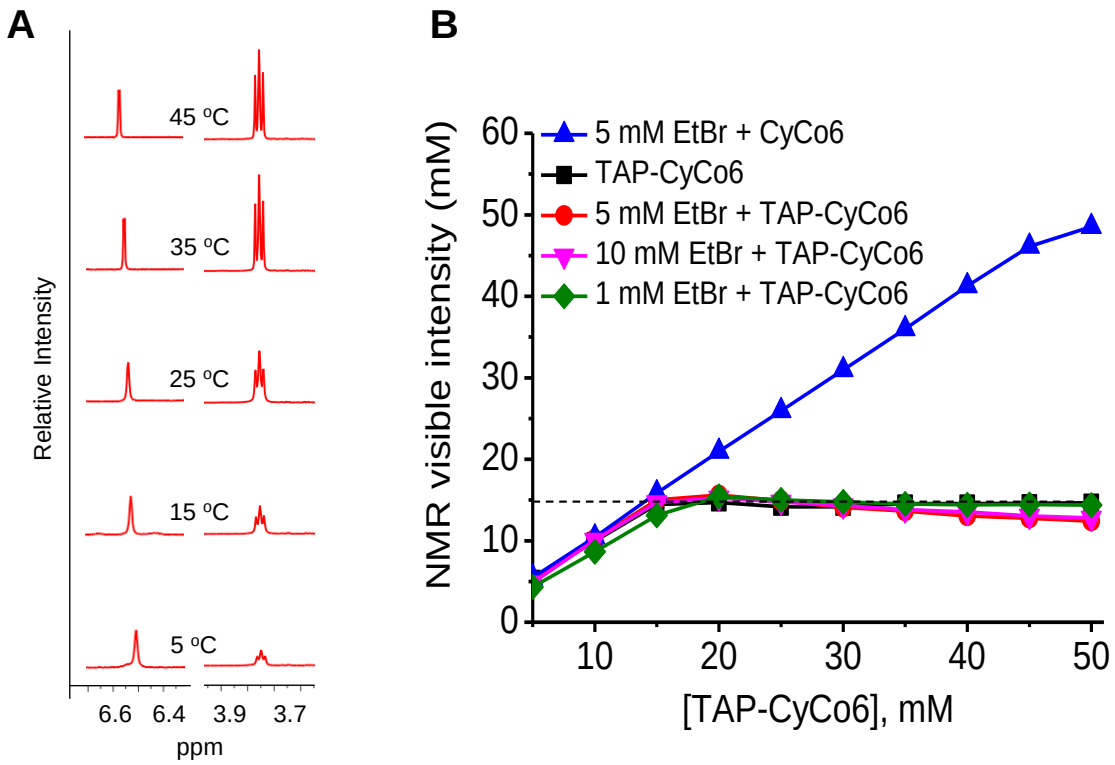


Figure 5.3 Ethidium bromide does not affect the minimal assembly concentration of TAP and CyCo6 . (A) Representative ^{1}H NMR spectra of 30 mM TAP-CyCo6 assembly in the presence of 1 mM EtBr at temperatures varying from 5 to 45 ^{o}C. Resonance at 6.5 ppm corresponds to an EtBr proton and resonance at 3.6 ppm belongs to a proton of CyCo6 (B) Plot of NMR-visible resonance intensity of unassembled CyCo6 *vs.* actual monomer concentration of both TAP and CyCo6 in solution with various concentrations of EtBr. Dashed line in plot represents the minimal assembly concentrations of TAP and CyCo6. Data collected by Dr. Suneesh Karunakaran.

5.3.3 Multiple Planar Polycyclic Molecules Inhibit Gelation

We next investigated nine additional planar polycyclic heterocycles (Figure 5.4) for their ability to alter hydrogel elasticity, as well as to more fully elucidate the structural properties that enable small molecules to control the length of CyCo6-TAP assemblies. Vial inversion tests were performed to evaluate the effect of 1 or 2.5 mM of each molecule on a solution of TAP and CyCo6 (30 mM each monomer). The heterocycles tested all had three or more aromatic rings and one or multiple charged moieties. As with EtBr, all of the molecules with positively charged moieties caused gel-to-solution transitions when incorporated at concentrations at or below of 1 mM, while no change in the gel strength was observed with the negatively charged molecules naphthol blue black and fluorescein (even up to 5 mM). This observation can be understood as positively charged molecules would more favorably interact with anionic periphery of the supramolecular assemblies prior to stacking with the rosette, while negatively charged molecules would be repelled. This molecular recognition event is similar to that which occurs between small molecules that intercalate nucleic acids (molecules which are all known to be positively charged) and the anionic phosphate backbone of these duplexes. The polycylic heterocycles with larger surfaces (i.e., those with more than 3 ring systems) were found to be more effective in transitioning the gel to a free-flowing solution. For instance, the addition of 0.05 mM cuprolinic blue to a gelled solution caused complete loss of the gel and qualitatively had a similar affect to solutions containing 1 mM EtBr.

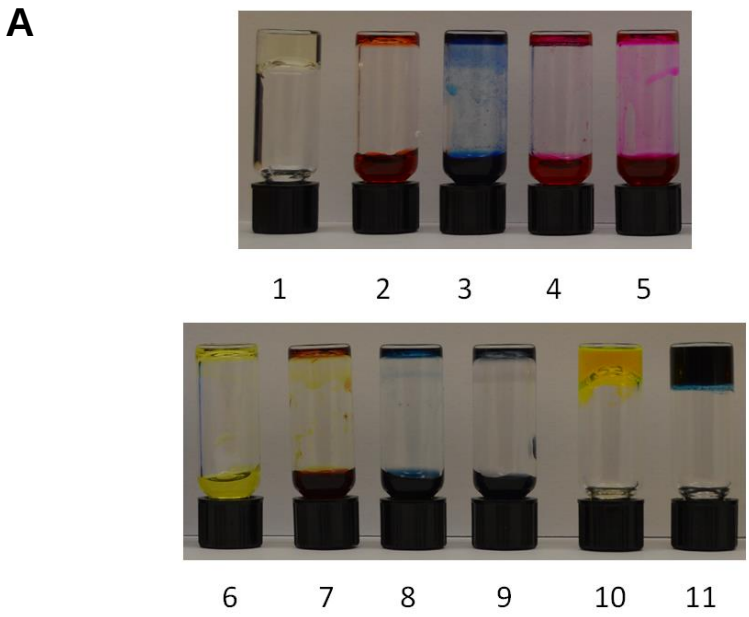


Figure 5.4 Multiple cationic planar polycyclic molecules cause a gel to solution transition in TAP-CyCo6 hydrogels. (A) Inverted bottle test containing TAP-CyCo6 assemblies with or without one of the polycyclic heterocycles screened in this study. (1) TAP-CyCO6 solution without additional plycyclic heterocycle, and with the addition of (2) ethidium bromide, (3) methylene blue, (4) rhodamine B, (5) Nile blue, (6) coralye chloride, (7) porphyrin (TMPyP), (8) cuprolinic blue (9) phthalocyanine (HoA), (10) fluorescein soidium, and (11) naphthol blue black. All solutions contained 30 mM of TAP and CyCo6 and 2.5 mM of various dye molecules when present. (B) The chemical structure of the polycylic heterocycles screened in this study. Data collected by Dr. Suneesh Karunakaran.

5.3.4 Tuning the Viscoelastic Properties of Hydrogels with Small Molecules

A more thorough analysis of the effect of the small molecules on the hydrogel's bulk mechanical properties was next performed using oscillatory rheology. Change in the dynamic moduli of solutions containing 30 mM TAP and CyCo6 in the absence and presence of various concentrations of EtBr was first performed (Figure 5.5). A shift in the crossover point, the intersection between G′ and G″, from lower to higher frequencies corresponds to a more fluid and less gel-like solution. With an increase in the concentration of EtBr within the CyCo6-TAP solutions from 0 to 1 mM, the crossover point shifted to larger frequencies with G″ becoming the dominating term over the course of the frequency sweeps. Additionally, with increasing EtBr concentration a decrease in both dynamic moduli occurred indicating a loss in both the viscosity and the elastic properties of these solutions. When porphyrin or cuprolinic blue, heterocycles with larger surface areas than EtBr, were tested in place of EtBr at 0.125 mM, the crossover point was found to shift to even larger frequencies (Figure 5.5). For example, compared to EtBr the effective concentration is approximately an order of magnitude lower at which porphyrin and cuprolinic blue impart similar dynamic moduli to the assembly solution. These results are in good agreement with the visual observation that increased flow of TAP-CyCo6 solutions is proportionate to both the surface area and amount of the polycyclic molecule in solution.

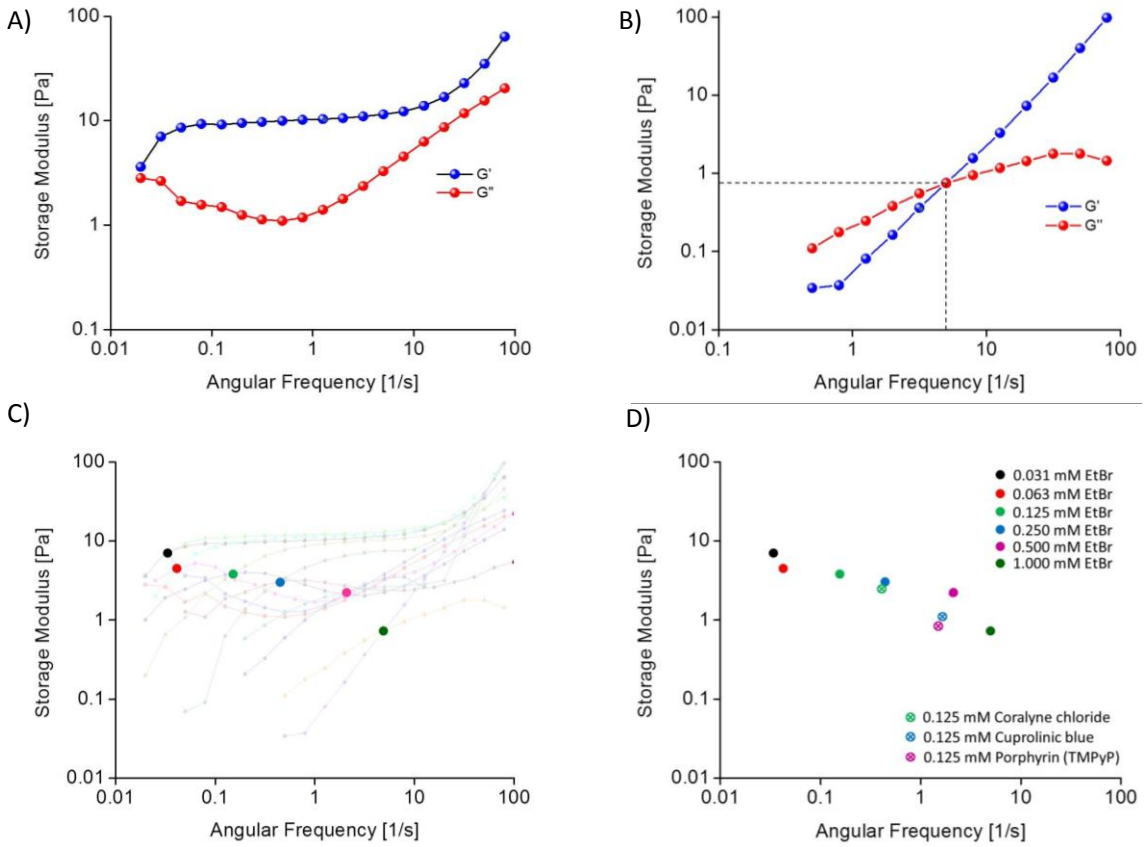


Figure 5.5 Small molecule dependent change in the dynamic shear storage modulus (G′) and loss modulus (G″) of solutions containing TAP-CyCo6 assemblies. (A and B) Solutions contained TAP and CyCo6 at 30 mM each monomer in the presence of (A) 0 mM and (B) 1 mM EtBr. (C) G′ and G″ and their crossover points at various EtBr concentrations. (D) Crossover points determined for solutions containing 0.125 mM coralyne, cuprolinic blue or porphyrin along with the crossover points shown in C for comparison with EtBr. All solutions contained 30 mM CyCo6 and TAP and 200 mM sodium phosphate buffer pH 7.Data collected by Dr. Miguel Pelaez.

Dynamic light scattering (DLS) measurements further supports the rheological data that bulk properties of CyCo6-TAP hydrogels can be tuned in both a concentration dependent and molecule specific fashion. Figure 5.6 shows the semi-logarithmic plots of the normalized autocorrelation functions of solutions containing EtBr, coralyne, porphyrin, and cuprolinic blue at concentrations ranging from 0.0034 mM to 2 mM. For all four molecules the correlation functions were found to change from broad and multi-exponential decays with high delay time (t) to approximately a single exponential profile with lower delay time. Concentrations above 2 mM EtBr were also investigated, however larger aggregates were observed and for this reason all studies were performed at or below 2 mM. These DLS results support the previous rheological data and further suggest that cuprolinic blue and porphyrin are the most affective noncovalent terminating agents that were investigated.

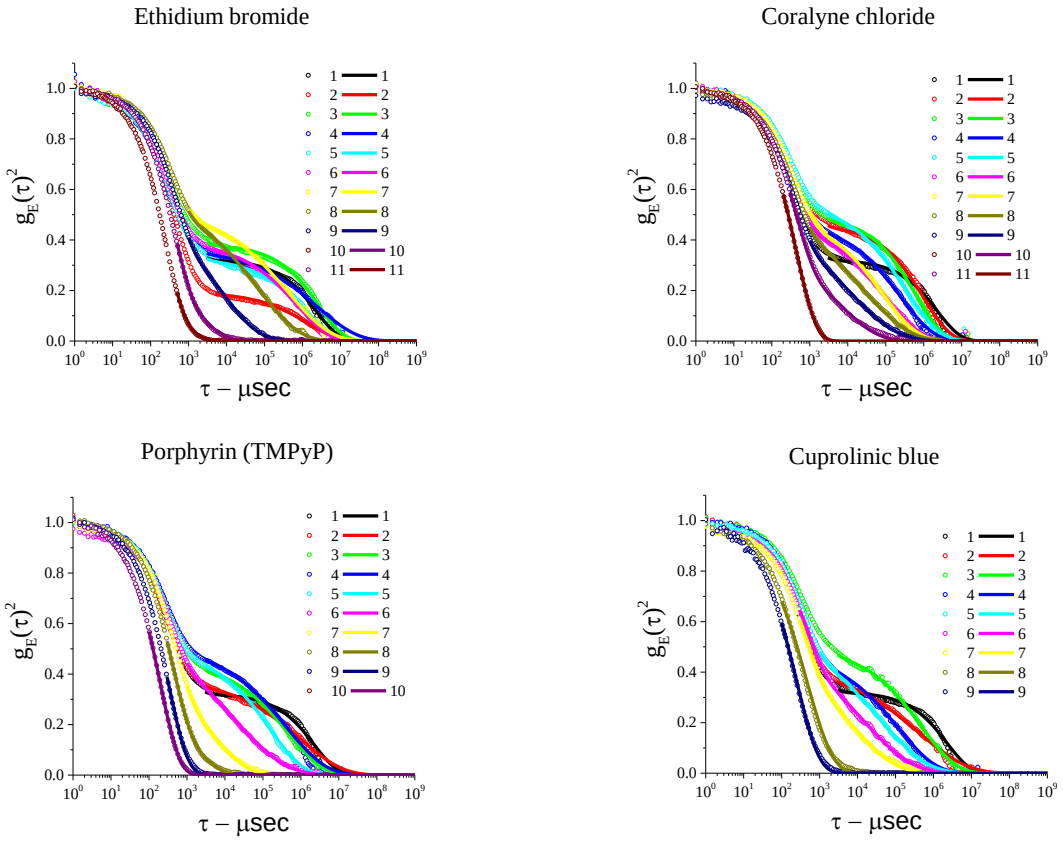


Figure 5.6 DLS measurements of solutions containing TAP-CyCo6 assemblies and various concentrations of candidate noncovalent terminators. Time correlation curves were generated from solutions containing TAP-CyCo6 assemblies and various concentrations of either EtBr, coralyne, porphyrin or cuprolinic blue. Concentrations ranged from 0.0034, shown as curve 1, to 2 mM shown as curve 10. All solutions contained 30 mM TAP and CyCo6. Data collected by Dr. Suneesh Karunakaran.

5.4 Conclusion

Results presented in this chapter demonstrate that polycyclic and planar heterocycles can efficiently control the length of supramolecular polymers. The proposed mechanism of action is through end-capping of the rosette surfaces at the termini of the supramolecular polymers by the small molecules, which we refer to as noncovalent terminators. As described in Chapter 2, long supramolecular polymers (greater than 5 microns) result from the tendency of the TAP-CyCo6 system to reduce the number of hydrophobic rosette surfaces exposed to water [16]. Frustration of polymer growth may result if a given noncovalent terminator can stack on the rosette surface and block stacking interactions between the surfaces of two rosettes. The noncovalent terminators that were found to be most effective contained larger surface areas that better approximated those of the TAP-Cy rosette. For example, cuprolinic blue (with a pthalocyanine stacking surface) altered the viscoelastic properties of the hydrogels at a 10-fold lower concentration than EtBr (with a phenanthridinium stacking surface). This finding is similar to the observation that planar polycyclic molecules that match the surface area of nucleic acid base-pair are better intercalators [17, 18]. It should be noted that even for the noncovalent terminators that interact most favorably with the rosette assembly (e.g., cuprolinic blue) the association constants for these molecule are greater than an order of magnitude smaller than the association constant between two rosettes; likely due to the fact that rosette surfaces are perfectly matched (which as described in the last chapter is most energetically favorable). Additionally, charge may be beneficial beyond providing solubility to the polycyclic heterocycle as the termini of

supramolecular polymer capped by a charged planar molecule would be expected to interact with water more favorably than the neutral rosette surface.

Additional experiments and data analysis are required for a thorough understanding of how noncovalent terminators control the degree of polymerization for the TAP-CyCo6 system. At present, the noncovalent binding of terminator molecules to supramolecular polymers via end-capping appears to modulate polymer length; a process that will depend on apparent association constants between the small molecule and the TAP-CyCo6 supramolecular assemblies. A greater understanding of this mechanism of controlling assembly length may provide new insight into methodologies to control the length of other supramolecular polymerization systems and aid in the development of functional materials.

5.5 References

1. Schaefer, C., et al., *Controlling the Cooperativity in the Supramolecular Polymerization of Ionic Discotic Amphiphiles via Electrostatic Screening.* ACS Macro Letters, 2012. **1**: p. 830-833.

2. De Greef, T.F., et al., *Supramolecular polymerization.* Chem Rev, 2009. **109**: p. 5687-5754.

3. Deng, R. and X. Liu, *Supramolecular polymers: Chain growth in control.* Nature Chem, 2015. **7**: p. 472-473.

4. van der Zwaag, D., T.F.A. de Greef, and E.W. Meijer, *Programmable Supramolecular Polymerizations.* Angew Chem Int Ed Engl, 2015. DOI: 10.1002/anie.201503104.

5. Whitesides, G.M. and B. Grzybowski, *Self-assembly at all scales.* Science, 2002. **295**: p. 2418-2421.

6. Lehn, J.-M., *Perspectives in Chemistry - Steps towards Complex Matter.* Angew Chem Int Ed, 2013. **52**: p. 2836-2850.

7. Hunter, C., *Supramolecular chemistry: Bigger and better synthesis.* Nature, 2011. **469**: p. 39-41.

8. Besenius, P., et al., *Controlling the growth and shape of chiral supramolecular polymers in water.* Proc Natl Acad Sci USA, 2010. **107**: p. 17888-17893.

9. Ogi, S., et al., *Living supramolecular polymerization realized through a biomimetic approach.* Nature Chem, 2014. **6**: p. 188-195.

10. Kumar, J., et al., *Self-discriminating termination of chiral supramolecular polymerization: tuning the length of nanofibers.* Angew Chem Int Ed Engl, 2015. **54**: p. 5943-5947.

11. Pal, A., et al., *Controlling the Structure and Length of Self-Synthesizing Supramolecular Polymers through Nucleated Growth and Disassembly.* Angew Chem Int Ed Engl, 2015.

12. Kang, J., et al., *Noncovalent assembly. A rational strategy for the realization of chain-growth supramolecular polymerization.* Science, 2015. **347**: p. 646-651.

13. Cafferty, B.J., et al., *Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK(a) matching of biomimetic monomers.* Chem Sci, 2014. **5**: p. 4681-4686.

14. Fetters, L.J., et al., *Connection between Polymer Molecular Weight, Density, Chain Dimensions, and Melt Viscoelastic Properties.* Macromolecules, 1994. **27**: p. 4639-4647.

15. Hirst, A.R., et al., *Low-molecular-weight gelators: elucidating the principles of gelation based on gelator solubility and a cooperative self-assembly model.* J Am Chem Soc, 2008. **130**: p. 9113-9121.

16. Cafferty, B.J., et al., *Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues.* J Am Chem Soc, 2013. **135**: p. 2447-2450.

17. Buckley, R., et al., *Molecular Recognition of Watson-Crick-Like Purine-Purine Base Pairs.* ChemBioChem, 2011. **12**: p. 2155-2158.

18. Persil, Ö., et al., *Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding.* J Am Chem Soc, 2004. **126**: p. 8644-8645.

CHAPTER 6

PREBIOTIC FORMATION OF TRIAMINOPYRIMIDINE NUCLEOSIDES THAT CO-ASSEMBLE WITH CYANURIC ACID INTO SUPRAMOLECULAR POLYMERS AND HYDROGELS[f]

6.1 Introduction

Forty years ago, Orgel and coworkers explored abiotic nucleoside formation by drying and heating D-ribose with each of the four canonical nucleobases (adenine, guanine, cytosine, and uracil). Only adenine was found to form nucleosides in detectable amounts (~2% yield) [1-4]. The poor glycosylation of the canonical nucleobases in model prebiotic reactions prompted the Orgel laboratory to explore the possibility that pyrimidine nucleosides were originally synthesized by nucleobase formation on a sugar scaffold [5], a hypothesis for which a complete synthetic pathway was more recently demonstrated by Sutherland and coworkers through an elegant, albeit multistep, reaction [6, 7]. Exploring the hypothesis that RNA evolved from a pre-RNA polymer with different nucleobases, Miller and coworkers [8], and later our laboratory [1], demonstrated that nucleosides are formed in good yields by reacting alternative heterocycles with ribose, specifically urazole and 2-pyrimidinone (Figure 6.1). These findings suggest that alternative nucleotides could have been more abundant on the prebiotic earth, or elsewhere in the solar system, than the canonical nucleosides.

f. This chapter was adapted from previously published work and is reproduced with permission. Chen, MC; Cafferty, BJ; Mamajanov, I; Gallego, I; Khanam, J; Krishnamurthy, R.; Hud, NV "Spontaneous Preboiric Formation of a β-Ribofuranoside That Self-Assembles with a Complementary Heterocycle." J. Am. Chem. Soc. (2014).

Even if a prebiotic route to all four canonical nucleosides were established, the mechanism by which these nucleosides could have been selected from a complex mixture of molecules and coupled into RNA polymers without the aid of enzymes remains a daunting question. A major challenge arises due to the fact that the canonical mononucleotides and their free bases do not form Watson-Crick base pairs in water [9, 10], but rather associate by π-stacking. Thus, it is difficult to imagine how the four nucleobases of contemporary RNA would have been exclusively selected from among similar, but non-pairing, heterocycles for incorporation into the first informational polymers of life. Hud and coworkers have proposed that this "paradox of base pairing" [11] could be overcome if there existed a set of recognition elements that are able to self-sort into non-covalent polymer assemblies prior to linkage by a common backbone [12]. Together, both prebiotic nucleoside formation and nucleobase selection may have not been so difficult if alternative heterocycles amenable to glycosylation and self-assembly formed the first informational polymers [13, 14].

In search of alternative nucleobases that could form nucleosides robustly and make stable base pairs in water, we systematically considered all possible pyrimidines and purines with $-NH_2$, =O, or -H as exocyclic groups. Of more than eighty possible candidates, 2,4,6-triaminopyrimidine (TAP) was one of the molecules chosen for investigation because its structure suggested the potential for enhanced reactivity with ribose [15], and because TAP is well known to form supramolecular assemblies with complementary heterocycles [16-20]. The H-bonding interactions between these heterocycles are similar to those found in Watson-Crick pairing; however, their ability to interact along three "faces" enables the formation of larger structures (see below).

Additionally, recent studies from our laboratory demonstrated the capability of modified, monomeric TAP to assemble with cyanuric acid (Cy) in aqueous solution to form micron-length supramolecular assemblies [17]. Here we show that aqueous mixtures of TAP and ribose spontaneously form β-furanosyl nucleosides in high yields. Upon the addition of the complementary Cy, the TAP-ribose conjugates assemble (within the unpurified reaction solution) to form micron-length polymers consisting of thousands of ordered nucleosides.

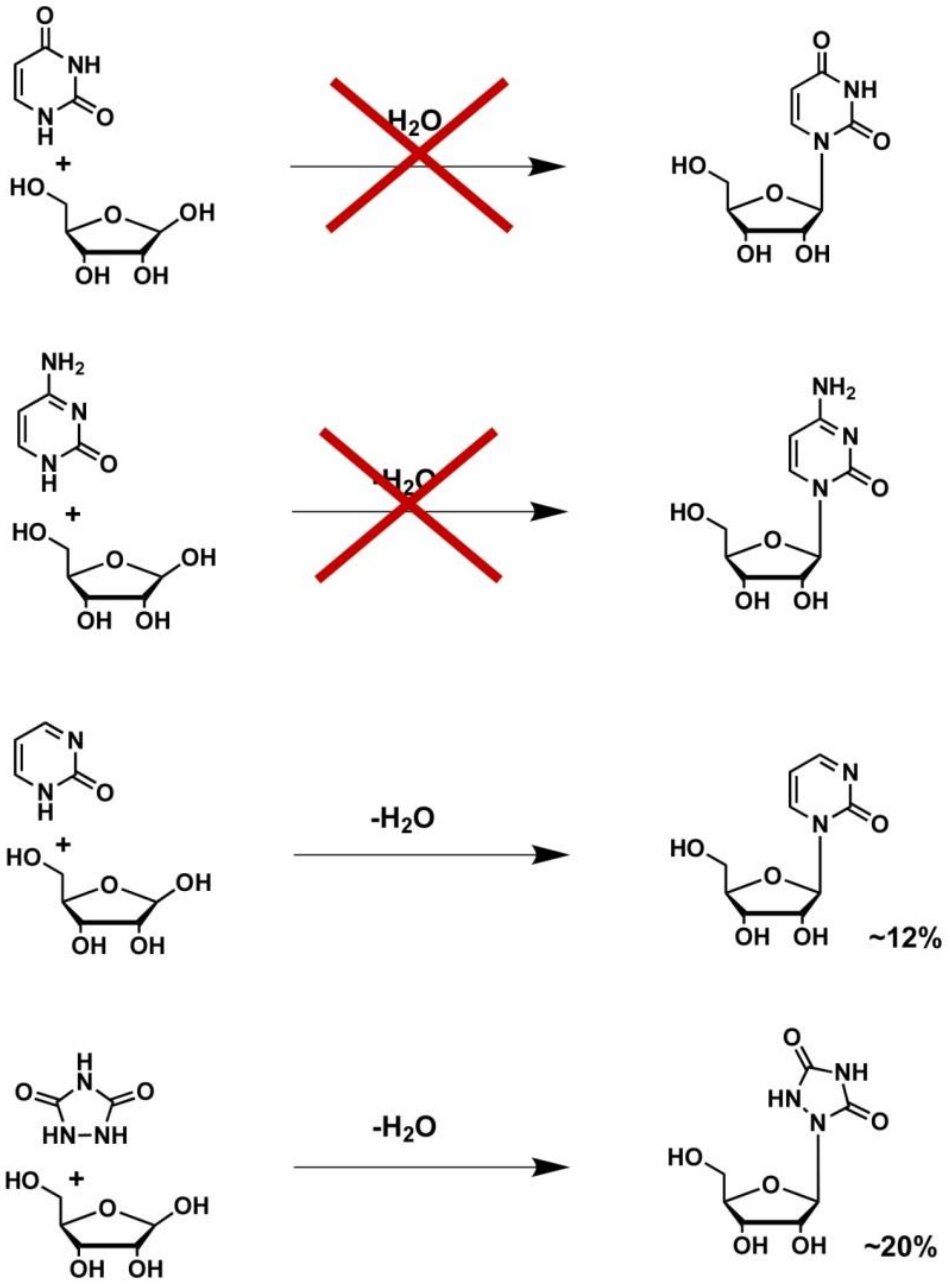


Figure 6.1 Nucleosidation reactions between ribose and select pyrimidines as well as the pyrimidine analog urazole.

6.2 Experimental Procedures

6.2.1 Materials

2,4,6-triaminopyrimidine (TAP) and cyanuric acid (Cy) was purchased from Acros Organic, D-ribose was purchased from either Amresco or Sigma-Aldrich. All materials were used without further purification.

6.2.2 Reactions of TAP with Ribose

Stock solutions of 2,4,6-triaminopyrimidine and D-ribose were prepared in nanopure H_2O. When dried at 45 to 95 °C, an Eppendorf tube with 1 mL of the stock solution was placed in a dry block heater at the specified temperature. For drying at 35 °C, tubes were placed in a vacuum centrifuge. Samples that were reacted for multiple days were dried for 24 hr, the resultant highly viscous solution was momentarily resuspended in the original volume of nanopure H_2O, and the process was repeated for the indicated number of days. Momentary hydration was performed to ensure that the solution was homogeneous during the course of the reaction. Additionally, daily hydration of the reactions mixture models environmental day/night cycles, a process which would have been operative on the prebiotic Earth. When reactions were performed in water without drying, TAP and ribose were combined at 300 mM and 600 mM, respectively, and samples were wither placed in an oven at 35 °C or 65 °C, or in a refrigerator at 4 °C. Analytical HPLC of samples was performed by running 95% 0.010 M NH_4HCO_3 (pH 7.2) and 5% CH_3OH at 0.5 $mL \cdot min^{-1}$ through a Phenomenex Synergi Polar-RP 80 Å column (250 mm × 4.60 mm × 4 µm).

6.2.3 Preparative Synthesis of 5 β-ribofuranosyl-2,4,6-triaminopyrimidine

2,4,6-triaminopyrimidine (1.5 mmol) and D-ribose (3 mmol / 2 eq.) were dissolved in 5 mL of nanopure H_2O. The sample was subjected to drying in a vacuum centrifuge for one week. The resulting light brown, highly viscous crude mixture was resuspended in 2.5 mL of 50 mM CH_3COONa buffer at pH 4.60. The solution was then loaded onto a gravity column containing Sephadex C50 Sulfopropyl cation-exchange media, and was eluted with a gradient of CH_3COONH_4 in 50 mM CH_3COONa buffer at pH 4.60. The resultant product-containing fractions were concentrated under vacuum to remove water. TARC was purified using an Agilent 1200 Infinity system by ion-pairing chromatography on a semi-preparative Phenomenex Prodigy ODS(3) 100 Å column (250 mm × 10 mm × 5 µm), with a mobile phase of 100% 0.1 M triethylammonium acetate. The resultant product-containing fractions were concentrated under vacuum, removing the majority of volatile salts from the various purification steps. Additional desalting was accomplished by reverse phase chromatography on the previously mentioned semi-preparative column. Maximum yield of TARC obtained was 20%.

UV/Vis (H_2O, pH 7): λ_{max} = 276 nm ε_{276} = 13,600 $mol \cdot L^{-1} \cdot cm^{-1}$. ^{1}H-NMR (500 MHz, D_2O) δ 4.57 (d, 9.1 Hz, H1'); 4.04 (dd, 9.1, 6.2 Hz, H2'); 3.92 (dd, 2.6, 6.2 Hz, H3'); 3.77 (m, 2.5 Hz, H4'); 3.57 (m, ABX, 2.0, 12.5 Hz, H5'a); 3.49 (m, ABX, 2.4, 12.6 Hz, H5'b). ^{13}C-NMR (126 MHz, D_2O) δ 161.93 (C4/C6); 160.6 (C2); 84.9 (C4'); 76.2 (C1'); 81.3 (C5); 70.9 (C2'); 70.6 (C3'); 60.4 (C5'). HRMS (pos. *m/z*): $C_9H_{16}O_4N_5$ theoretical mass: 258.1197, actual mass: 258.1199.

6.2.4 Crude Reaction and TARC Gel Preparation

The crude gel was formed by combining Cy (1.5 mmol / 1 eq.) with the 10 day dried TAP + ribose reaction product (see preparative synthesis above) and then dissolved in 300 mM sodium borate (pH 8) to reconstitute the sample at 5 ml. Reactions containing Cy were carried out the same as with the TAP+ribose reactions except 1 equivalent of Cy was included with the starting material. In all cases the samples were then heated to 95 °C for 5 min to dissolve the Cy and cooled at room temperature. A cloudy, yellow hydrogel was observed with an off white precipitate collected at the bottom. Samples from the original crude reaction, precipitate, and hydrogel were collected and analyzed by HPLC-DAD. Co-injection and LCMS analysis were used to identify TAP from TAP-ribose conjugates. Clarified hydrogels were obtained by centrifugation of the crude gel through a 0.2 micron Millipore spin column. Gels prepared with purified TARC were made by the same procedure with the addition of 200 mM sodium phosphate to buffer at pH 7.

6.2.5 Analytical Techniques

CD spectra were obtained on a Jasco J-720 CD spectrometer, and acquired by scanning 200-350 nm at a rate of 500 nm/min, with averaging of twenty measurements. UV absorption analysis was performed on an Agilent 8453 UV-vis spectrophotometer equipped with an 89090A temperature controller. All samples that were inspected by CD or UV contained 200 mM sodium phosphate and 200 mM boric acid (pH 7) and were analyzed in a strain-free sandwich cell with either a 0.1 or 0.01 mm path length. The structure of TARC was then confirmed by NMR using ^{1}H-NMR, ^{13}C-NMR, COSY,

HSQC, HMBC, and ROE, all collected on a Bruker DRX-500. High-resolution mass spectrometry was performed on a Waters Synapt G2. Purity was confirmed by analytical HPLC as described in the analytical synthesis of TARC.

Samples were imaged by AFM over a silicon surface. Prior to deposition, silicon wafers were rinsed with nanopure water followed by ethanol and dried under a N_2 stream. After rinsing, silicon surfaces were cleaned with UV-ozone for 30 minutes (JE- LIGHT UVO Cleaner Model #42) and stored in a petri dish until use. A 1.5-µl sample that was stored on ice was spin coated (Laurell Technologies) at 4 ºC for 30 sec (2000 rpm, or 4000 rpm). After spin coating, the sample was dried at room temperature in a Petri dish overnight. The AFM imaging was performed on a Multimode AFM with Nanoscope IIIa controller and a "J" scanner (Veeco Instruments) in tapping mode in air, using Si tips (Vistaprobes, 40 N/m).

6.3 Results and Discussion

6.3.1 Drying and Heating of TAP with Ribose Produces TAP-Ribose Conjugates

TAP and D-ribose were mixed together in water and heated under various conditions to assess the glycosylation potential of TAP. As shown in Figure 6.2, a considerable fraction of TAP becomes glycosylated to form TAP-ribose conjugates when the mixtures were dried at 35 °C. The reaction is robust, with TAP-ribose conjugates being produced in comparable yields when the ratio of TAP and ribose were varied over a factor of five, and when the reaction was carried out at temperatures ranging from 35 to 95 °C. The yield of single-ribose TAP conjugates was maximized relative to double- and triple-ribose additions to TAP when mixtures were dried at the lower range of temperatures investigated and with TAP:ribose ratios close to unity. As shown in Figure 6.1B, HPLC analysis with simultaneous MS and UV absorption monitoring reveals that vacuum-assisted drying of a 1:2 TAP:ribose mixture at 35 °C for one day results in over 30% of TAP being modified by conjugation with ribose. Extended reaction times allowed maximum conversion of TAP to TAP-ribose conjugates. For example, the yield of single-ribose TAP conjugates (including nucleosides, see below) increased to around 60% after ten days of drying at 35 °C, with momentary daily rehydration (which allowed the reaction to be carried out for multiple days under conditions of low water activity while remaining homogenous) (Figure 6.2C). Reactions carried out at higher temperatures and with higher ratios of ribose to TAP resulted in the more rapid and nearly complete modification of TAP, with up to 90% of all TAP being modified by the addition of one to three ribose molecules. This reaction was observed to happen even in aqueous solution (i.e., no drying), within days between 35 to 65 °C, or weeks at 4 °C (Figure 6.3).

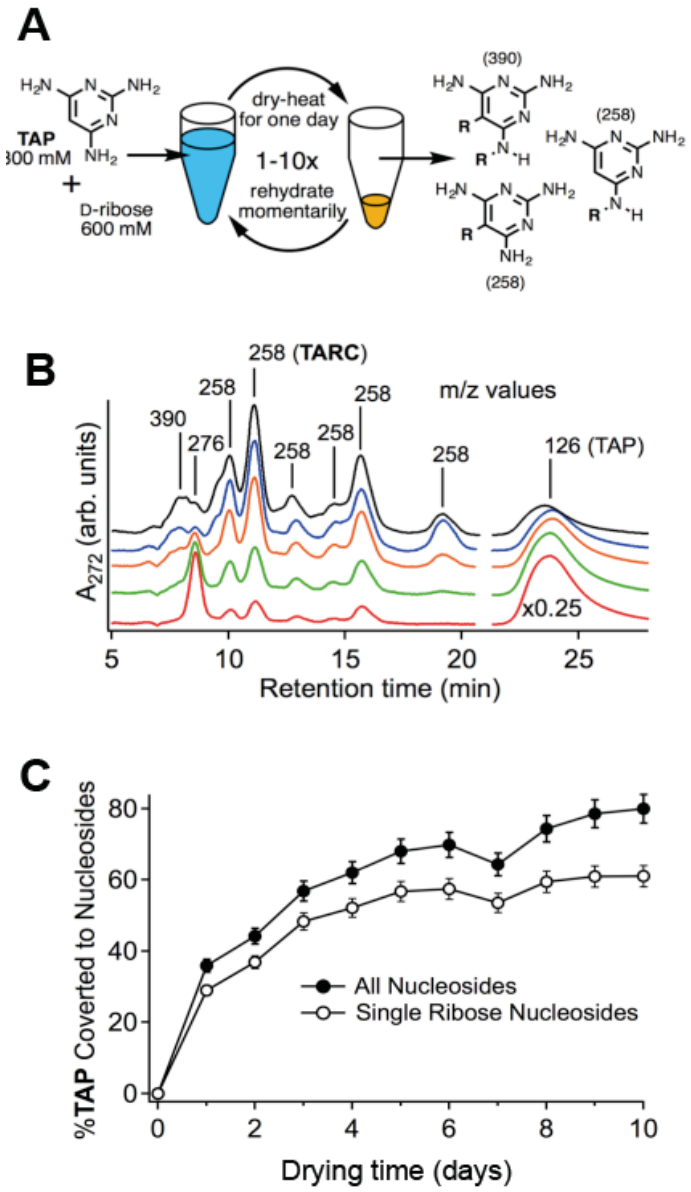


Figure 6.2 Formation of TAP+ribose conjugates by dry-heating reactions. (A) Chemical structure of 2,4,6-triaminopyrimidine (TAP) and the process used to generate TAP-(D)-ribose conjugates (including nucleosides). R groups on TAP represent any possible ribose isomers. (B) HPLC chromatographs of samples of TAP and ribose after one to five days of drying, i.e., drying of TAP (300 mM) with ribose (600 mM) (pH 8) at 35 °C (vacuum assisted) with momentary rehydration every 24 hr. HPLC traces, from bottom to top, correspond to reaction times of 1, 2, 3, 4, and 5 days. Peaks are labeled with m/z values obtained by simultaneous MS and UV absorption monitoring of LC. The m/z values listed correspond to TAP with one closed-ring ribose conjugate, 258; TAP with two conjugated riboses, 390; and TAP conjugated with an open chain ribose, 276. (C) Plot of percentage of TAP converted to nucleosides as a function of reaction time (same reaction conditions as in (B). Error bars represent known sources of experimental uncertainty.

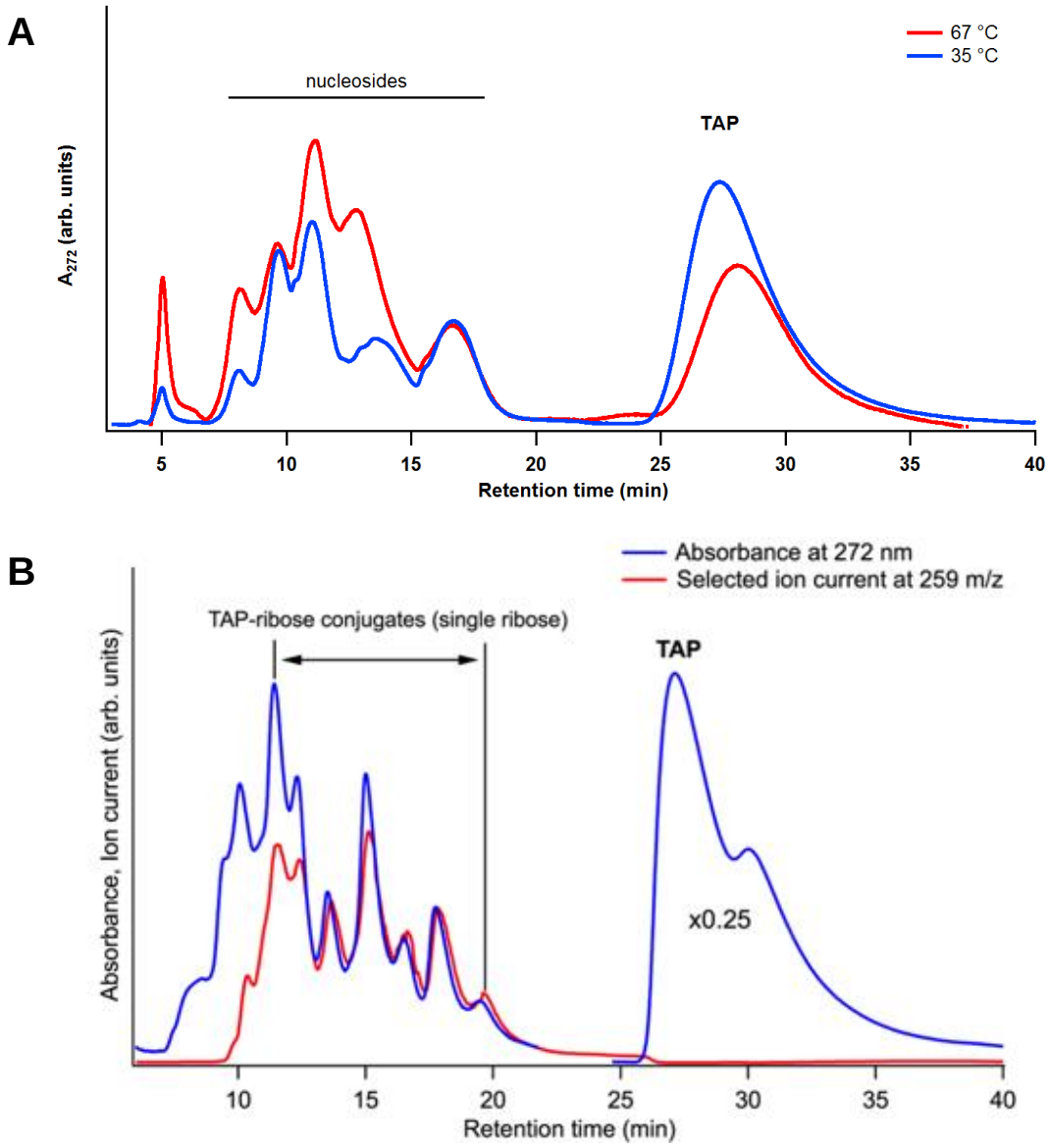


Figure 6.3 Formation of TAP+ribose conjugates in bulk water. (A) LC-UV analysis of TAP-ribose conjugates formed spontaneously by TAP and ribose in water after reaction for 4 days at either 35 °C or 65 °C. (B) LC-UV and LCMS analysis of TAP-ribose conjugates formed spontaneously in water at 4 °C for 105 days. Chromatogram monitored by UV absorbance at 272 nm (blue curve) illustrates that approximately 20% of TAP has been modified by the addition of one or more ribose molecules. Simultaneous monitoring of MS during LC separation revealed that the major species for peaks with retention times between 11.5 and 19.5 min (marked with vertical lines) correspond to TAP-ribose nucleosides with single ribose additions (i.e., 258 m/z). Additionally, the selective ion chromatogram for species with 258 m/z (red curve) illustrates that clear majority of species with absorbance at 272 nm and retention times between 11.5 and 19.5 min are TAP-ribose conjugates with single ribose additions. For both reactions, TAP (300 mM) and D-ribose (600 mM)

Drying TAP with ribose results in a thick syrup with free TAP and TAP-ribose conjugates reaching a concentration of ~38% (m/m). The solubility of TAP in the absence of ribose is only 3.6% (m/m), indicating that a substantial fraction of TAP forms associations with ribose, which could include highly labile linkages between open-chain ribose and the exocyclic amines of TAP that may not be observed during HPLC analysis. The increased solubility of TAP in the presence of ribose contrasts with the relatively low solubility of the canonical nucleobases under similar conditions. For example, cytosine (the most soluble of the four RNA nucleobases) has a solubility limit in water of 0.73% (m/m), and this limit does not appear to increase with ribose as a cosolute. In addition to increasing TAP solubility, ribose conjugation with TAP could also prevent ribose degradation by fixing the sugar in its cyclized form, a possible rudimentary mechanism to aid the accumulation of ribose that is complimentary to previously proposed prebiotic mechanisms, which include ribose complex formation with borate and sequestration by selective crystallization [21-23].

6.3.2 The Major Product of the TAP+ribose Reaction is a *β*-Ribofuranooside

To confirm that nucleosides are among the products of the TAP+ribose reaction, the main product (~20%) resulting from the ten-day 35 °C drying reaction of a 1:2 TAP+ribose mixture was isolated by HPLC. NMR spectroscopy revealed that this product is 5-*β*-ribofuranosyl-2,4,6-triaminopyrimidine (TARC) (Figures 6.4-6.6). 2D HMBC analysis confirmed that TARC is a *C*-nucleoside, with a C-C bond between ribose and the nucleobase (Figure 6.5), while 1D ROE experiments established TARC to be of the *β*-furanose conformation (Figure 6.6).

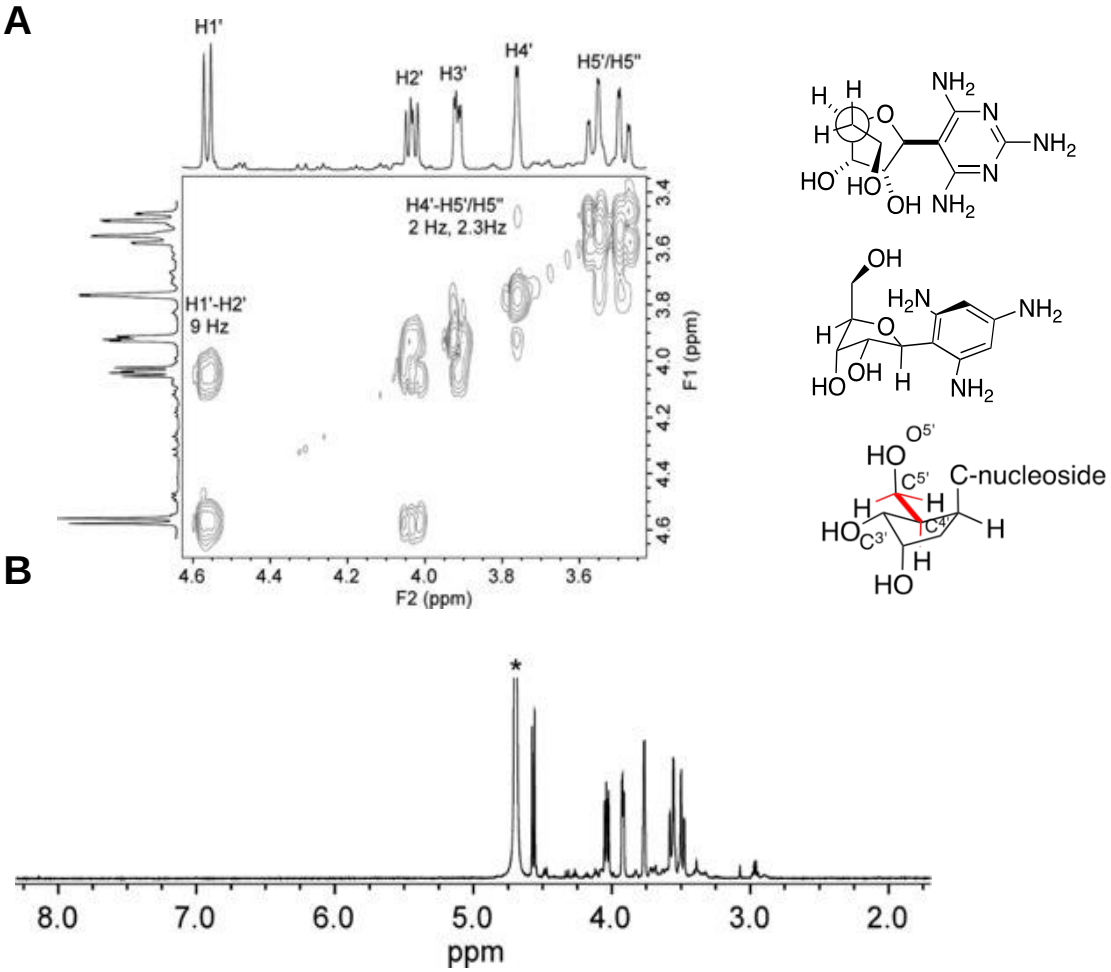


Figure 6.4 ^{1}H and 2D COSY analysis of TARC. (A) The 2D spectrum, and proton splitting patterns, used to confirm sugar proton assignments. (B) No resonances were observed in the aromatic region, which is supportive of a *C*-nucleoside chemical structure. The splitting of the H1' proton is sometimes used for *N*-nucleoside structure determination. However, the H1' proton couplings in *C*-nucleosides can be different from *N*-nucleosides. As observed by Chaudhuri et al. [24], the H1' proton of a *C*-nucleoside exhibits splitting of around 6.5 Hz for the α-furanose form and around 10.6 Hz for the β-furanose form. The TARC H1' splitting of 9 Hz is consistent with the β-furanose structure assignment. The relatively small H4'-H5'/H5" couplings (estimated at 2 and 2.3 Hz) is also supportive of the furanose sugar conformation (as in C), as the pyranose form is expected to exhibit one strong (e.g., ~10 Hz) and one weak H4'-H5'/H5" coupling. Additional data supporting the *C*-β-furanosylribose structural assignment of TARC is presented below. Peak marked with * in B is HDO solvent resonance. c) The relatively small H4'-H5'/H5" couplings are suggestive of the torsion angle (O5'–C5'–C4'–C3') that is concordant with synclinal arrangement (irrespective of the sugar pucker).

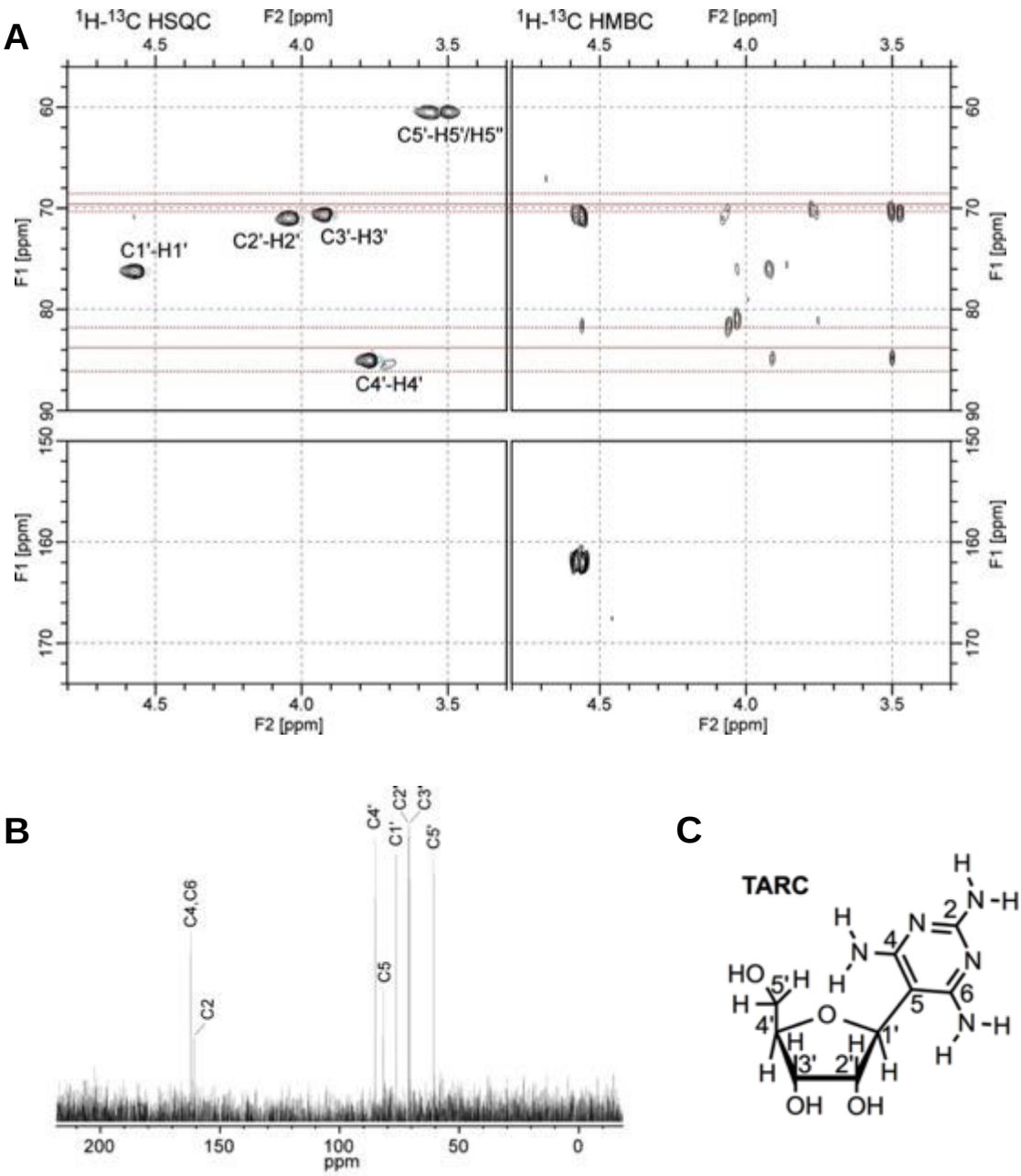
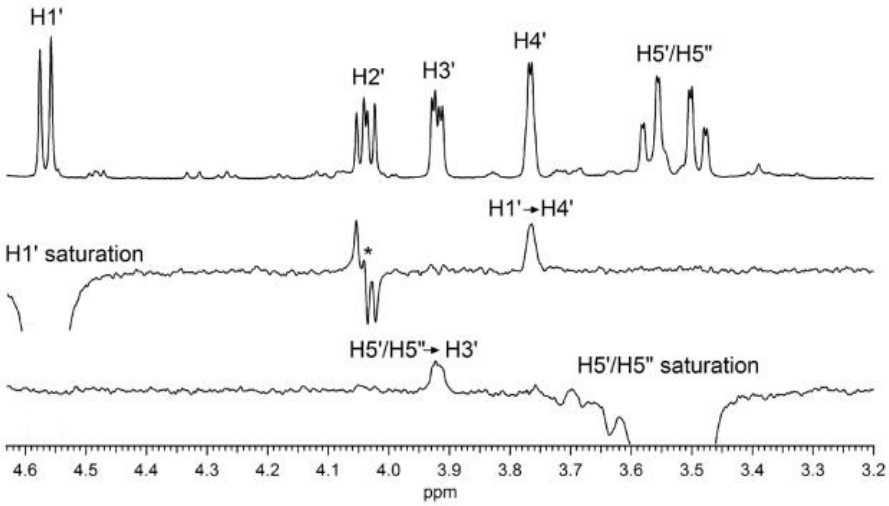


Figure 6.5 HSQC and HMBC analysis of TARC. (A) The HMBC spectrum was optimized for 10 Hz ^{1}H-^{13}C couplings. The C4' chemical shift can be used differentiate between pyranose and furanose ribonucleosides. For example, the C4' chemical shifts compiled by Beier and Mundy [25] for the model nucleoside methyl- β -ribopyranose ranged from 68.6 to 70.4 ppm, with a mean value of 69.6 ppm (indicated by upper set of dotted and solid red lines); chemical shifts for methyl-β-ribofuranose ranged 81.9 to 86.1 ppm, with a mean value of 83.7 ppm (indicated by the lower set of dotted and solid red lines). Based upon these previous assignments, the observed C4' chemical shift of 85.6 ppm is indicative of TARC having ribose in the furanose form. (B) Full ^{13}C spectrum of TARC. (C) TARC structure with carbon numbering.

A



B

Theoretical distances (Å)	H1'/H3' (r13)	H1'/H4' (r14)	*(r14/r13)^6*	H1'/H5'	H5'/H2'	H5'/H3'
β-ribofuranoside	3.8	2.5	*0.08*	4.3	3.5-5.0	2.6-3.0
α-ribofuranoside	3.6	3.7	*1.2*	4.2-4.5	4.0-5.0	2.6-3.0
β-ribopyranoside	3.8	4.0	*1.4*	2.3	4.1	3.9
α-ribopyranoside	4.0	3.8	*0.74*	3.6	4.0	3.8

Figure 6.6 ROEs and theoretical atomic distances in TARC. (A) ^{1}H NMR and 1D ROE spectra of the nonexchangeable (ribose) protons of TARC. (Top) ^{1}H NMR spectrum with resonance assignments. (Middle) Irradiation of the H1′ results in through-space magnetization transfer to the H4′ proton. *indicates TOCSY transfer from H1' to H2', a through-bond magnetization transfer by these strongly coupled protons (e.g., 9Hz). (Bottom) Irradiation of the H5'/H5" protons results in through-space magnetization transfer to the H3' proton. (B). To aid assignment of sugar conformation using ROE results, a table is presented with theoretical through-space distances between H1′ and H5′/H5″ protons with other ribose protons that are diagnostic of nucleoside sugar structure. Because the H5' and H5" protons are irradiated together in the ROE experiment, the shortest distance for the two protons is given under label H5'. For the furanosides the two values provided for H5′ protons correspond to distances associated with the two lowest energy rotamers about the C4′-C5′ bond. Note that the observation of a H1'->H4' ROE as the strongest through-space magnitization transfer is indicative of a *β*-furanose structure, as this nucleoside structure is the only one out of the four possible structures that the H1'->H4' distance as the shortest of the proton-proton distances for H1' and H5'. All spectra were collected at 5 °C in D_2O. Distances given in the table for the four different possible TARC *C*-nucleoside isomers are based on measurements from energy-minimized, Chem3D, ACD System.

Contemporary nucleosides are predominantly *N*-ribosides, but *C*-nucleosides do exist in nature, including pseudouridine, the most common RNA post-transcriptional nucleoside modification [26]. The finding that the ribose of TARC is in the β-furanose conformation, like the nucleosides found in extant life, was somewhat unexpected as free ribose exists predominantly in its β-pyranose form [27]. This discovery is consistent with previously reported model prebiotic reactions of nucleoside formation, which have also indicated that the ribofuranosyl form of ribose could have been available since the earliest abiotic glycosylation reactions [1, 8].

Like other 2,4,6-substituted pyrimidines, which undergo Mannich reactions [28], TAP is known to be nucleophilic at the C5 position [29]. TARC formation is expected to proceed through the reaction of TAP with the free aldehyde form of ribose [30]. This proposed pathway, detailed in Figure 6.7, is supported by the observation of a covalent intermediate in early wet-dry cycles with the mass of an open chain ribose connected to TAP (peak with m/z of 276 in Figure 6.2B). Consistent with this putative assignment, the 276 m/z product is a transient species, appearing most prominently after the first two days of drying, but vanishing almost completely after the fifth day of drying. In contrast, TARC yield steadily increases up to at least five days of drying and does not decrease with additional drying time. It is also a certainty that exocyclic *N*-glycosides are among the TAP-ribose conjugates present in the crude reaction mixture, as multiple products with m/z values corresponding to TAP with one, two, and three covalent ribose additions are observed (which necessitates that some of these products have at least one ribose added to an exocyclic amine of TAP). Furthermore, formation of exocyclic *N*-glycosides involving the natural nucleobases has previously been shown to form in both drying

reactions [2, 3, 5] and in solution [31]. Hydrolysis studies of the crude TAP-ribose reaction mixture show the rapid conversion of multiple ribose-nucleosides to free TAP or a single-ribose TAP conjugate (Figure 6.8), which is analogous to the facile release of ribose from the exocyclic amine of adenine [2, 3].

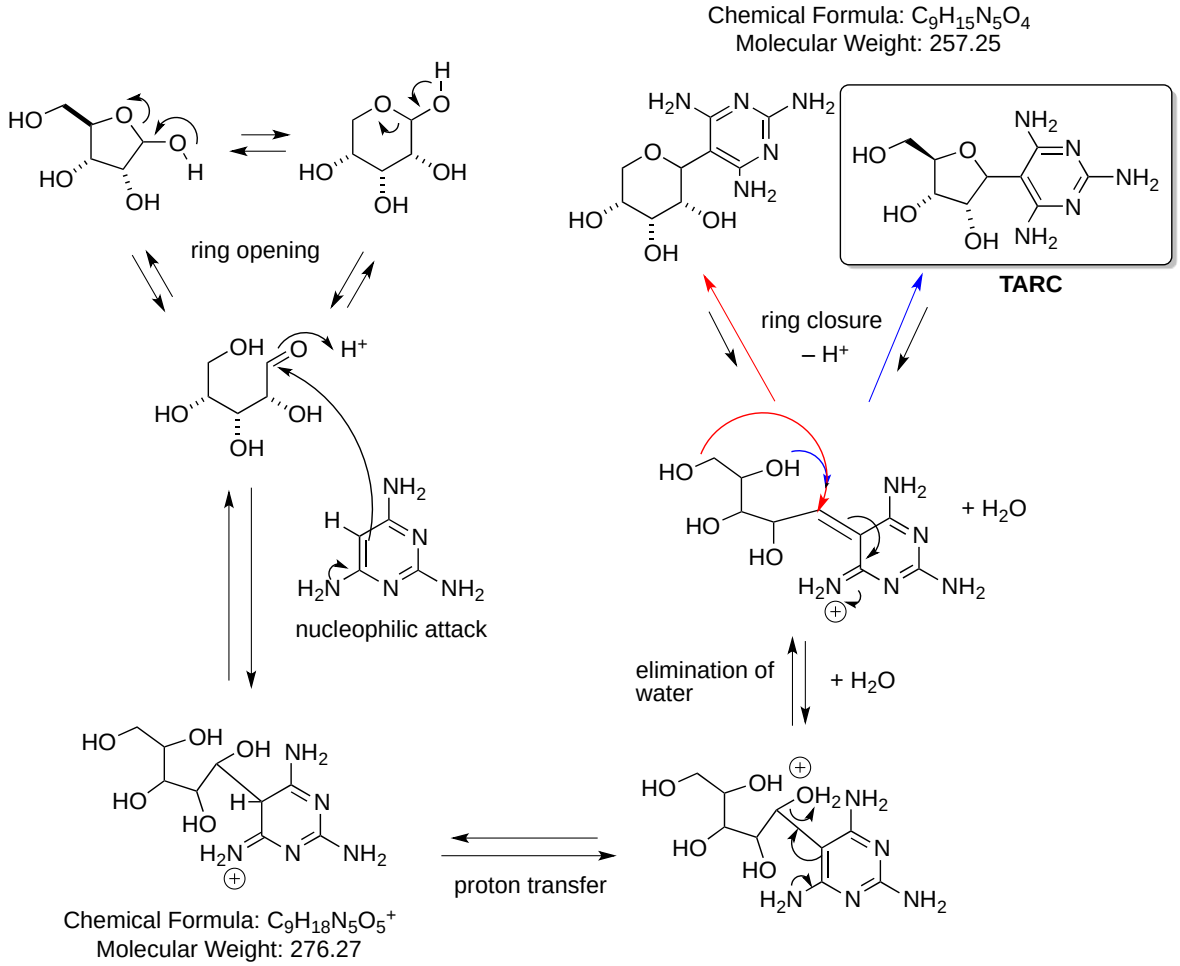


Figure 6.7. Proposed mechanism for TARC formation from ribose and TAP. The presence of 2,4,6-triamino groups increases the electron density at C(5) position, which attacks the aldehydic carbon of the open form of ribose. Following elimination of water, the conjugated iminium intermediate is attacked, either by the attack of the C(5′)-OH or the C(6′)-OH group, leading to ring closure forming the *C*-nucleosides. Overall, this is a ribosylation of the TAP (or nucleosidation of ribose by TAP) by an electrophilic aromatic substitution mechanism. Mechanism proposed by Prof. R. Krishnamurthy.

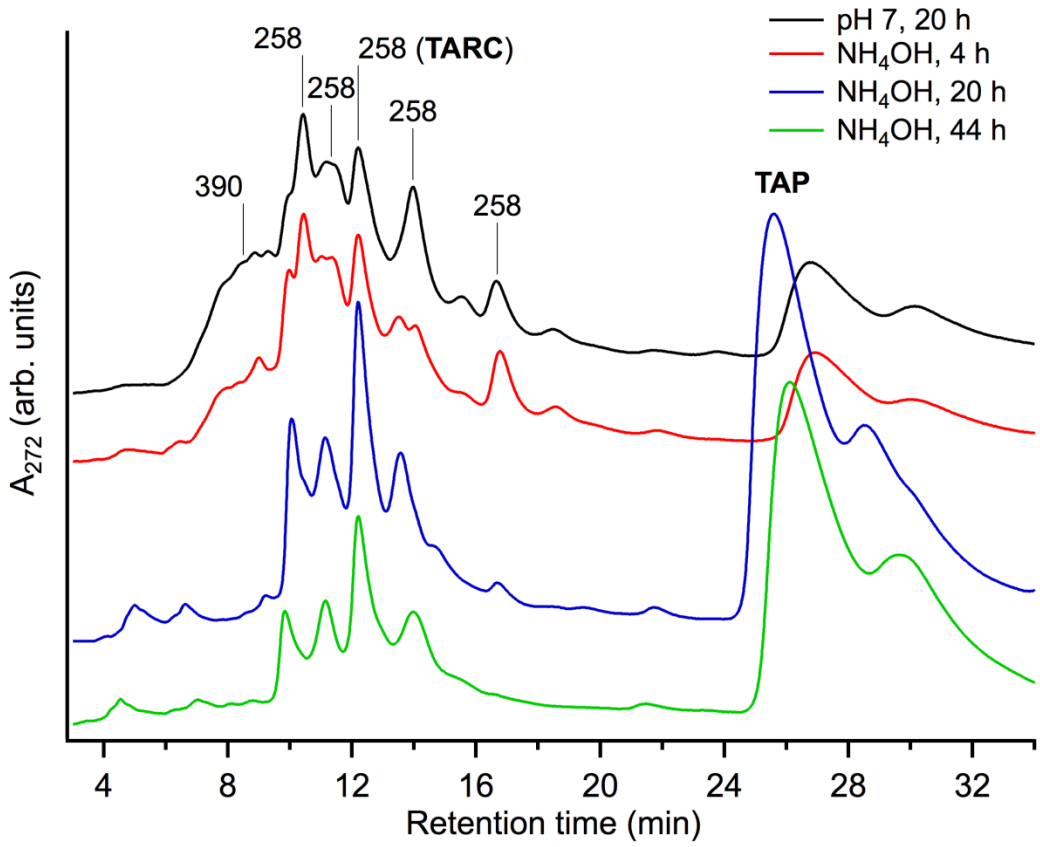


Figure 6.8 LCMS analysis of base-catalyzed selective degradation of TAP nucleosides. The crude reaction products of TAP and ribose (dried at 35°C for ten days; top, black) were exposed to 10 M ammonium hydroxide at 65 °C for 4 to 44 hr. Double additions to ribose (390 m/z), as monitored by LCMS, decrease with increased time of incubation with ammonium hydroxide. Note that four dominant peaks remain. This result is consistent with the four possible *C*-nucleosides (i.e., α-furanose, β-furanose (TARC), α-pyranose, β-pyranose) being most stable to hydrolysis.

6.3.3 TARC Assembles with Cy to Form a Hydrogel with Micron-Length Polymers

HPLC purified TARC was tested to determine if it is able to assemble with cyanuric acid (Cy) in aqueous solution. Mixing TARC with Cy in equimolar proportions in a sodium phosphate and boric acid buffer resulted in the formation of a shear-thinning hydrogel (Figure 6.9). Hydrogel formation is indicative of self-assembly into supramolecular polymers that further associate to form a continuous matrix [32].

Circular dichroism (CD) was used to inspect solutions containing TARC and Cy, as dramatic changes in CD-signal intensity are often associated with the formation of supramolecular assemblies by chiral monomers [33, 34]. In the present case, large, positive bands from 210-250 and 270-290 nm were observed for TARC and Cy mixtures at 5 °C for concentrations greater than 5 mM in each monomer; a TARC and Cy mixture at 2.5 mM each exhibited a CD profile essentially identical to TARC alone (Figure 6.9B). Heating of a sample containing 5 mM TARC and Cy to 20 °C resulted in a complete loss of the assembly-induced CD signal. The TARC UV absorption spectrum exhibited a red-shift upon the addition of Cy, demonstrating the formation of J-type aggregates, a spectral feature that is associated with π-π stacking interactions (Figure 6.9C).

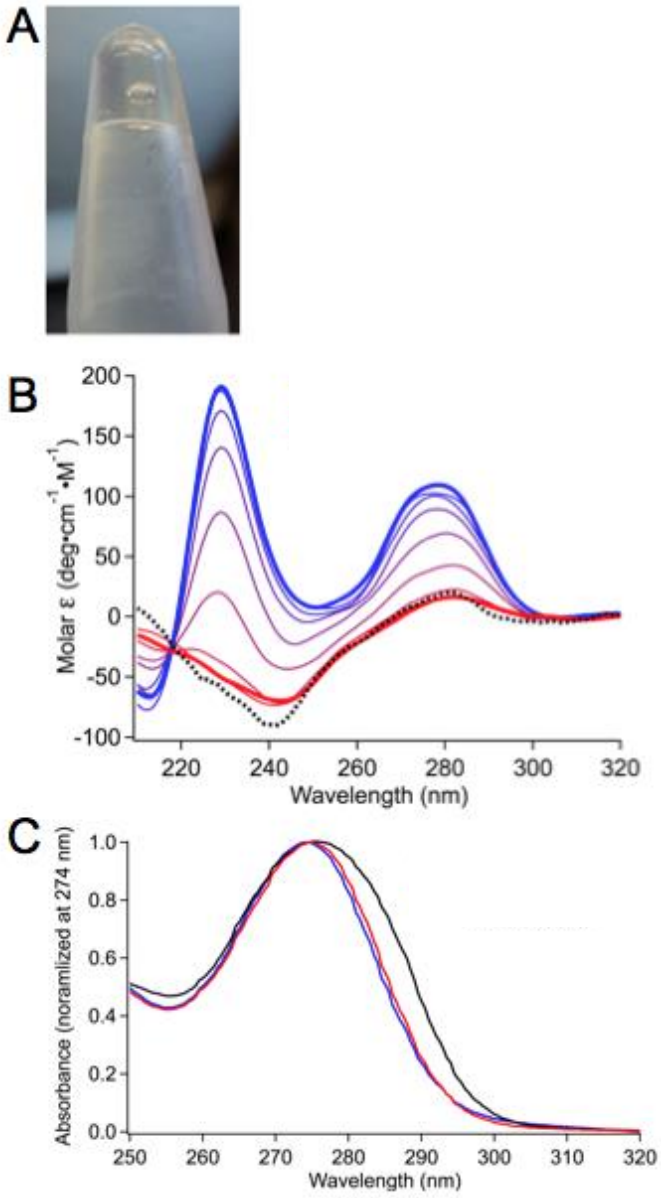


Figure 6.9 Gel formation and spectroscopic analysis of TARC+Cy assemblies. (A) Gel formed by TARC and Cy (40 mM in each monomer). High viscosity of the gelled solution is illustrated by bubble that remained in place indefinitely. (B) CD spectra of 5 mM TARC and Cy, in 200 mM sodium phosphate/200 mM sodium borate (pH 7), at temperatures ranging from 4 (blue trace) to 20 °C (red trace), in steps of 2 °C. Spectrum of TARC in same buffer without Cy is also shown for comparison (dotted line). (C) The UV spectrum of TARC (56 mM) and Cy (56 mM) at 4 °C in sodium phosphate buffer (200 mM, pH 7.0) with sodium borate (200 mM) in a 0.01 mm pathlength cuvette was taken (black), showing a red-shift characteristic of J-type aggregates when compared to TARC alone (blue). When the same solution was dilute by 100-fold (red), the TARC/Cy UV profile matches that of TARC alone. All absorbances were normalized to absorbance maximum within region shown.

TARC-Cy assemblies were visualized by atomic force microscopy (AFM), which revealed supramolecular polymers with lengths greater than 200 nm (Figure 6.10A). The diameter of a single fiber, measured perpendicular to the image plane, was 2 nm (insert, Figure 6.10A). Self-assembled structures of comparable thickness and length were recently observed when a molecule named TAPAS (TAP with succinate conjugated to an exocyclic amine) was mixed with Cy [17]. The similarities of these structures indicate that the fibers shown in Figure 6.10A are TARC nucleosides assembled with Cy in a similar manner to that of TAPAS assembled with Cy. That is, Cy and TARC are arranged within "rosettes" that are π-stacked to form linear supramolecular assemblies (Figure 6.10B). These assemblies have a predicted width of approximately 1.6 nm for TAP before conjugation with ribose, which is consistent with the measured fiber thickness of 2 nm. Given that the fibers imaged by AFM are up to 500 nm in length, and that the interplane stacking of planar ring systems is 0.34 nm, each of these linear assemblies represents up to 1500 stacked hexads.

The borate anion in the sodium borate buffer was found to be essential for the formation of water-soluble assemblies. Initial mixing experiments with TARC and Cy in the absence of borate resulted in immediate precipitation, as observed when unmodified TAP is mixed with Cy. The ability for borate to improve the solubility of the TARC-Cy complex was not unexpected. Water-soluble rosette assemblies by TAP (or melamine, its triazine analog) and Cy have only been obtained when a charged group is incorporated at the periphery of these assemblies [17, 35], and it was anticipated that the borate anion would add a negative charge to TARC upon complex formation with the vicinal *cis*-diols of ribose [22, 36].

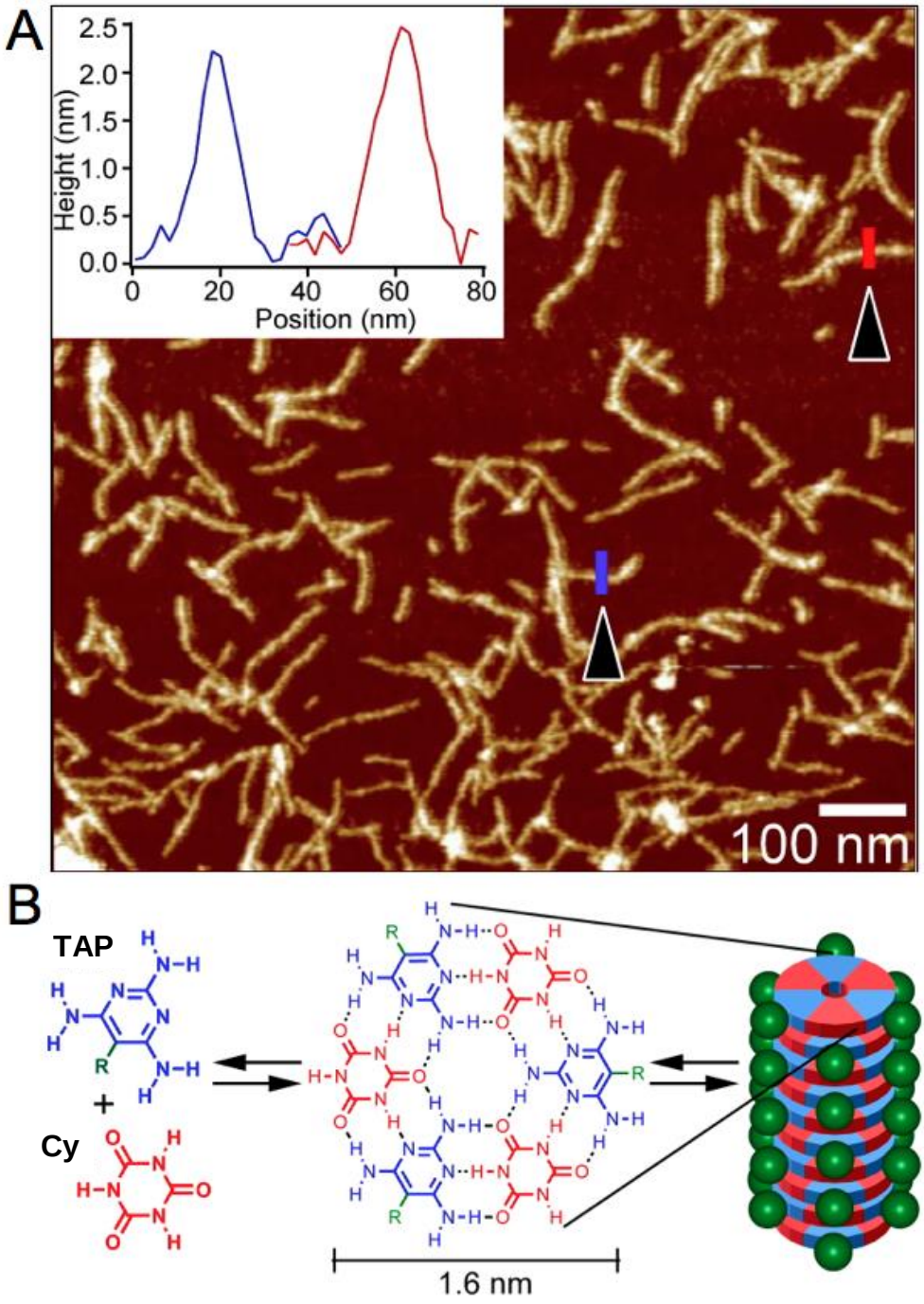


Figure 6.10 AFM of TARC-Cy assemblies. (A) AFM topographic image of self-assembled TARC-Cy fibers from a solution containing 50 mM in each molecule. Inset shows the height profile of single fibers delineated by the red and blue lines in the main panel. (B) Structures of TARC and Cy, their H-bonded rosette structure, and proposed higher-order assemblies based upon AFM fiber dimensions. Ribose is indicated as R on the TAP chemical structures and green spheres on schematic models of TAP.

6.3.4 TAP-Ribose Conjugates Assemble with Cy in the Crude Reaction Mixture

In order to find a more realistic one-pot prebiotic reaction for the formation *and selection* of nucleosides, we investigated the potential of TAP nucleosides in the reaction mixture to associate with Cy while still in solution with unreacted starting materials, non-pairing TAP-ribose conjugates, and any other reaction products. For this investigation, the crude reaction mixture of 1 part TAP and 2 parts ribose that had been subjected to drying for ten days was resuspended in sodium borate buffer, pH 8. Cy was then added to the solution and heated to 95 °C (to facilitate dissolving Cy), upon cooling to room temperature the solution formed a hydrogel that coexisted with a precipitate (Figure 6.11A). Control experiments in which TAP and ribose were freshly mixed without drying, or where Cy was not added to the post-drying TAP+ribose reaction mixture, did not result in hydrogel formation, confirming that the product(s) of the TAP+ribose reaction and their assembly with Cy are the cause of hydrogel formation. Similar hydrogels were also obtained if Cy was initially added to an unreacted solution containing TAP and ribose, followed by drying and heating; confirming that the order of Cy addition is not critical to the TAP+ribose reaction or hydrogel formation (Figure 6.11B).

The translucent hydrogel solution formed by the crude TAP+ribose reaction mixture and Cy could be clarified of precipitate by centrifugation to yield a transparent hydrogel (Figure 6.11C). HPLC-UV analyses of the material in the gel and the precipitate revealed that the hydrogel is the preferential phase for TAP-ribose conjugates, whereas the precipitate is the preferential phase for unmodified TAP (Figure 6.11D). Specifically, the crude TAP+ribose gel contained approximately 87% of TAP as TAP-ribose

conjugates and 13% as free TAP. In contrast, the precipitate formed upon the addition of Cy was substantially depleted of TAP-ribose conjugates (43%) and enriched in free TAP (57%). We propose that this separation of TAP-ribose conjugates from free TAP by supramolecular polymer formation/precipitation with Cy represents a potential prebiotic mechanism to locally concentrate pairing nucleosides away from non-pairing molecules (e.g., non-pairing TAP-ribose conjugates) and from precipitating molecules that do pair (e.g., free TAP) in a complex chemical mixture.

The hydrogel phase formed by the crude TAP+ribose reaction mixture and Cy was imaged by AFM, again revealing the formation of micron-length fibers (Figure 6.12). The thickness of single fibers was determined to be 2 nm (insert, Figure 6.12), as found with purified TARC assembled with Cy. While TARC-Cy combination is likely to be a large constituent of these non-covalent assemblies, other TAP-ribose conjugates (as in Figure 2A) may also be part of these assemblies. Additional investigations are required to determine how differences in the TAP-ribose conjugates, such as sugar structure and conformation, may affect their selection and inclusion into the supramolecular assemblies. Regardless, the ability of the TAP+ribose conjugates and Cy to assemble into ordered supramolecular structures without the need for starting from a pure source of monomer demonstrates a robust and potentially prebiotic pathway for monomer isolation, which may allow for a pathway for 'self-selection' by a stepwise constitutional preferential exclusion/inclusion mechanism.

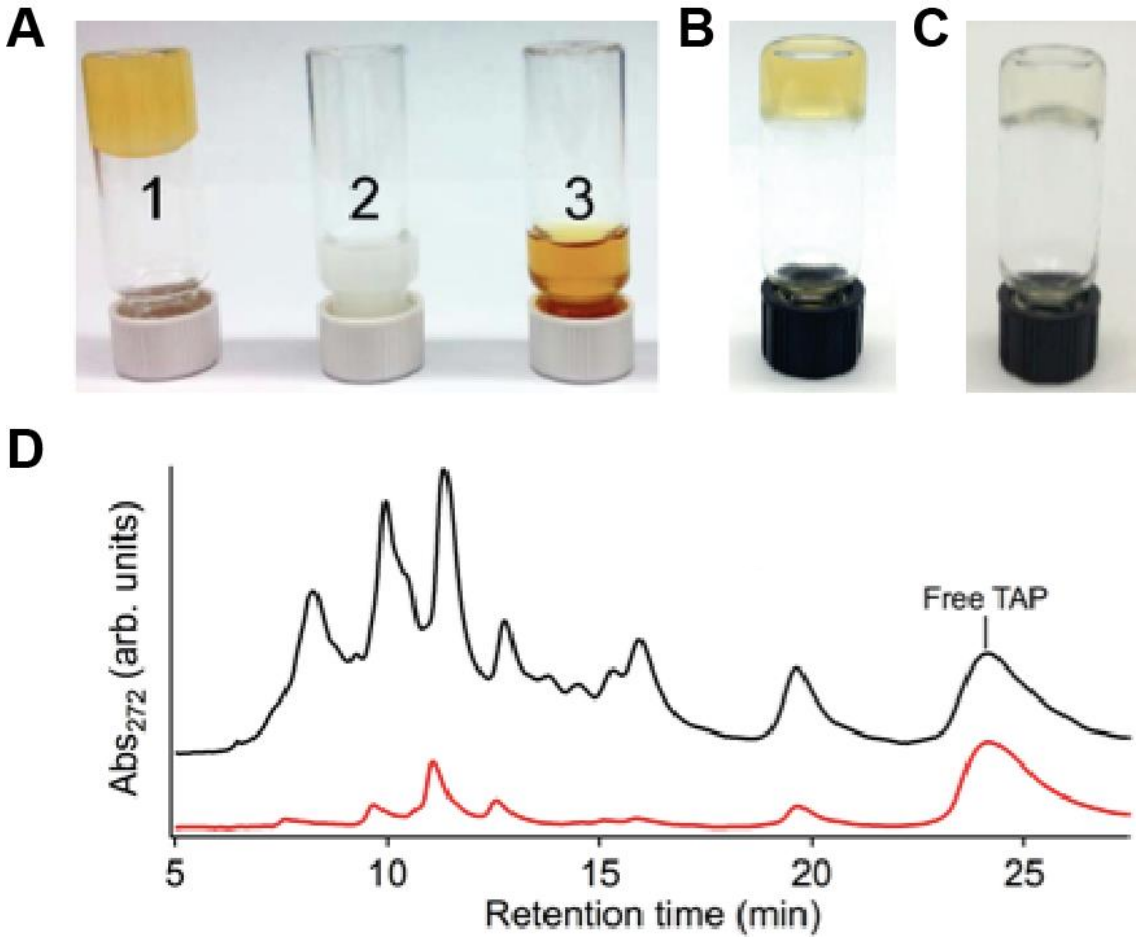


Figure 6.11 Inverted bottle test and HPLC analysis showing assembly in crude reaction solutions. (A) Inverted bottle test showing gel formation of a ten-day reaction of TAP with ribose (1:2::TAP:ribose; vacuum-assisted drying at 35 °C with daily rehydration) after addition of cyanuric acid (Cy) to the same concentration as TAP (modified and unmodified) (bottle 1). Cy was added with sodium borate (300 mM, pH 8) at room temperature. Gel formation is not observed if procedure is identical to that of bottle 1, but TAP and ribose solution is not subjected to drying (bottle 2), or if Cy is not added to sample (bottle 3). (B) Same as bottle 1 in A, except Cy was added before sample was dried to generate TAP-ribose conjugates. (C) Same as bottle 1 in A, except solution was clarified of precipitated material by centrifugation and diluted to 16 mM for both Cy and TAP containing species. (D) HPLC chromatographs of the crude TAP+ribose reaction product used to make the gel shown in C (upper trace), and of the precipitate that formed when this crude mixture was mixed with Cy (lower trace). Chromatographs were normalized to integrated area of free TAP peaks.

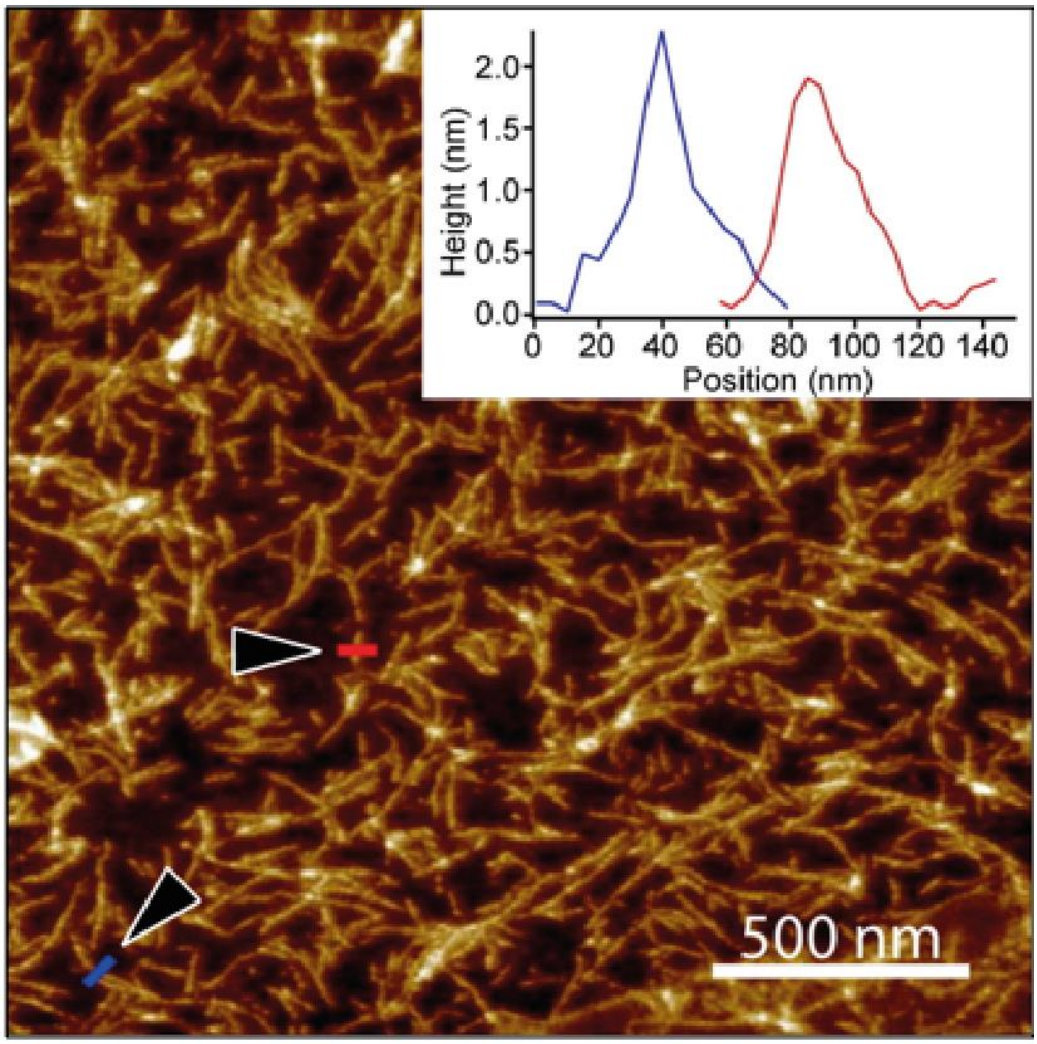


Figure 6.12 AFM topographical image of assemblies formed by the crude TAP+ribose reaction mixed with Cy. Insert shows height measurements across individual fibers indicated by red and blue lines and arrowheads in main image.

6.4 Conclusion

In summary, we have shown that TAP readily forms nucleosides with ribose in a one-pot model prebiotic reaction. The high yield of the β-ribofuranoside, TARC, from the TAP+ribose reaction contrasts with the low or non-existent nucleoside formation found in ribose reactions with the canonical nucleobases. These observations lend support to the hypothesis that the extant nucleosides may be products of evolution and that there could have been alternative nucleobases that were more amenable to prebiotic nucleoside formation [12, 37, 38]. The addition of a complementary H-bonding heterocycle with TAP nucleosides, even in the crude TAP+ribose reaction mixture, results in the spontaneous assembly of micron-length non-covalent polymers. Through an appropriate linking chemistry, these assemblies of ordered nucleosides would facilitate monomer coupling and thereby the *de novo* formation of covalent, proto-biopolymers. These results begin to address how alternative nucleobases that readily self-assemble could have set the stage for the emergence of the early informational polymers of life by both facilitating nucleoside formation and the selection of pairing bases from complex mixtures, before the emergence of enzymes.

6.5 References

1. Bean, H.D., et al., *Formation of a β-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction.* J Am Chem Soc, 2007. **129**: p. 9556-9557

2. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis. VI. synthesis of purine nucleosides.* J Mol Biol, 1972. **67**: p. 25-33.

3. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis: VII. solid-state synthesis of purine nucleosides.* J Mol Evol, 1972. **1**: p. 249-257.

4. Maurel, M.-C. and O. Convert, *Chemical structure of a prebiotic analog of adenosine.* Orig Life Evol Biosph, 1990. **20**: p. 43-48.

5. Sanchez, R.A. and L.E. Orgel, *Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides.* J Mol Biol, 1970. **47**: p. 531-543.

6. Powner, M.W., B. Gerland, and J.D. Sutherland, *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions.* Nature, 2009. **459**: p. 239-242.

7. Powner, M.W., J.D. Sutherland, and J.W. Szostak, *The origins of nucleotides.* Synlett, 2011. **14**: p. 1956-1964.

8. Kolb, V.M., J.P. Dworkin, and S.L. Miller, *Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides.* J Mol Evol, 1994. **38**: p. 549-557.

9. Solie, T.N. and J.A. Schellman, *The interaction of nucleosides in aqueous solution.* J Mol Biol, 1968. **33**: p. 61-77.

10. Ts'o, P., I. Melvin, and A. Olson, *Interaction and association of bases and nucleosides in aqueous solutions.* J Am Chem Soc, 1963. **85**: p. 1289-1296.

11. Engelhart, A.E. and N.V. Hud, *Primitive genetic polymers.* Cold Spring Harbor Perspectives in Biology, 2010. **2**. DOI: 10.1101/cshperspect.a002196.

12. Hud, N.V., et al., *The origin of RNA and 'My Grandfather's Axe'.* Chem Biol, 2013. **20**: p. 466-474.

13. Joyce, G.F., et al., *The case for an ancestral genetic system involving simple analogs of the nucleotides.* Proc Natl Acad Sci USA, 1987. **84**: p. 4398-4402.

14. Orgel, L.E., *Prebiotic chemistry and the origin of the RNA world.* Crit Rev Mol Biol, 2004. **39**: p. 99-123.

15. Rembold, H. and H.J. Schramm, *Kondensation des 2.4-diamino-6-hydroxy-pyrimidinsmit aldosen.* Chem Ber, 1963. **96**: p. 2786-2797.

16. Bohanon, T.M., et al., *Barbituric-acid 2,4,6-triaminopyrimidine aggregates in water and their competitive interaction with a monolayer of barbituric-acid lipids at the gas-water interface.* Angew Chem Int Ed, 1995. **34**: p. 58-60.

17. Cafferty, B.J., et al., *Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues.* J Am Chem Soc, 2013. **135**: p. 2447-2450.

18. Lehn, J.M., et al., *Molecular ribbons from molecular recognition directed self-assembly of self-complementary molecular-components.* Journal of the Chemical Society-Perkin Transactions 2, 1992: p. 461-467.

19. Marchi-Artzner, V., et al., *Molecular Recognition between 2,4,6-Triaminopyrimidine Lipid Monolayers and Complementary Barbituric Molecules at the Air/Water Interface: Effects of Hydrophilic Spacer, Ionic Strength, and pH.* Langmuir, 1998. **14**: p. 5164-5171.

20. Rakotondradany, F., et al., *Hydrogen-bond self-assembly of DNA-analogues into hexameric rosettes.* Chem Comm, 2005: p. 5441-5443.

21. Prieur, B.E., *Study on the potential probiotic activity of boric acid.* C. R. Acad Sci Paris, Chimie / Chemistry, 2001. **4**: p. 667-670.

22. Ricardo, A., et al., *Borate minerals stabilize ribose.* Science, 2004. **303**: p. 196-196.

23. Springsteen, G. and G.F. Joyce, *Selective derivatization and sequestration of ribose from a prebiotic mix.* J Am Chem Soc, 2004. **126**: p. 9578-9583.

24. Chaudhuri, N.C., R.X.F. Ren, and E.T. Kool, *C-nucleosides derived from simple aromatic hydrocarbons.* Synlett, 1997: p. 341.

25. Beier, R.C. and B.P. Mundy, *Assignment of Anomeric Configuration and Identification of Carbohydrate Residues by13C NMR: Arabino- and Ribopyranosides and Furancsides.* J Carbohydr Chem, 2006. **3**: p. 253-266.

26. Charette, M. and M.W. Gray, *Pseudouridine in RNA: What, Where, How, and Why.* IUBMB Life, 2000. **49**: p. 341-351.

27. Angyal, S. and V. Pickles, *Equilibria between pyranoses and furanoses. II. Aldoses.* Aust J Chem, 1972. **25**: p. 1695-1710.

28. Han, B., et al., *Mannich-type C-nucleosidations with 7-carba-purines and 4-aminopyrimidines.* Synlett, 2005. **16**: p. 0744-0750.

29. Delia, T.J., D.D. Kirt, and S.M. Sami, *An application of the mannich reaction using hydroxylamine and some of its derivatives - 6-substituted-2,4-diamino-5,6,7,8-tetrahydropyrimido 4,5-D pyrimidines.* J Heterocyclic Chem, 1983. **20**: p. 145-147.

30. Dworkin, J.P. and S.L. Miller, *A kinetic estimate of the free aldehyde content of aldoses.* Carbohydr Res, 2000. **39**: p. 359-365.

31. Johnson, K.M., et al., *On the Formation and Properties of Interstrand DNA–DNA Cross-Links Forged by Reaction of an Abasic Site with the Opposing Guanine Residue of 5′-CAp Sequences in Duplex DNA.* J Am Chem Soc, 2012. **135**: p. 1015-1025.

32. Estoff, L.A.H., A. D., *Water gelation by small organic molecules.* Chem Rev, 2004. **104,**: p. 1201-1217.

33. Fenniri, H., et al., *Helical rosette nanotubes: Design, self-assembly, and characterization.* J Am Chem Soc, 2001. **123**: p. 3854-3855.

34. Hirschberg, J.H.K.K., et al., *Helical self-assembled polymers from cooperative stacking of hydrogen-bonded pairs.* Nature, 2000. **407**: p. 167-170.

35. Ma, M.M. and D. Bong, *Determinants of cyanuric acid and melamine assembly in water.* Langmuir, 2011. **27**: p. 8841-8853.

36. Springsteen, G. and B. Wang, *A detailed examination of boronic acid–diol complexation.* Tetrahedron, 2002. **58**: p. 5291-5300.

37. Rios, A.C. and Y. Tor, *Refining the genetic alphabet: a late-period selection pressure?* Astrobiology, 2012. **12**: p. 884-891.

38. Rios, A.C. and Y. Tor, *On the Origin of the Canonical Nucleobases: An Assessment of Selection Pressures across Chemical and Early Biological Evolution.* Isr J Chem, 2013. **53**: p. 469-483.

CHAPTER 7

SPONTANEOUS FORMATION AND MOLECULAR ASSEMBLY OF HYPOTHETICAL ANCESTRAL NUCLEOTIDES IN WATER

7.1 Introduction

Ultimately underpinning the RNA World hypothesis is the formation of complementary nucleotides through abiotic processes. As discussed in the preceding chapters, adenosine has been formed in 1-5% yield when solutions of adenine and ribose are dried and heated [1, 2] and cytosine and uracil have been synthesized from small molecule precursors through a multistep water-based synthesis [3]. While encouraging, plausible prebiotic reactions have not been reported that produce complementary (i.e., pairing) nucleosides – a property that was likely essential for the emergence of RNA-based life [4, 5]. Persistent difficulties associated with model prebiotic synthesis of the canonical nucleosides (e.g., low yields, multiple steps, required intermediate purification), led to the proposal that RNA has evolved from an earlier *proto*-RNA polymer through chemical and/or biological evolution [6, 7]. Demonstrations of ribose glycosylation by prebiotically accessible heterocycles [2, 8, 9] indicate that nucleotides with nucleobases slightly different from those of extant RNA may have been common on the prebiotic Earth. As discussed in the last chapter, 2,4,6-triaminopyrimidine (TAP) will react with ribose to form nucleosides in good yield and assemble with cyanuric acid (Cy) in water [10]. Nevertheless, the TAP heterocycle has not yet been reported in a model prebiotic reaction and Cy appears too unreactive to form glycosidic linkages with ribose under model prebiotic conditions.

As part of our search for the possible components of proto-RNA, we evaluated a series of heterocycles for the ability to spontaneously form nucleosides with ribose *and* form stable base-paired assemblies in water [10]. Here we report nucleosidation reactions and nucleotide assembly/mutual selection of two candidate proto-nucleobases, melamine (MA) and barbituric acid (BA) (Figure 7.1). These heterocycles are ideally suited to function as the recognition units of a primitive genetic system, as both MA and BA: i) have been produced in the same model prebiotic reactions [11]; ii) can form Watson-Crick-like base pairs with each other [12]; iii) have H-bond donor and acceptor groups that are complementary with either uracil or adenine (for MA and BA, respectively), making these heterocycles forward compatible for base pairing with two of the extant nucleobases (Figure 7.1) [4]; iv) possess chemical properties that indicate favorable glycosylation by ribose, specifically at the C5 position of BA and the exocyclic amines of MA. Both MA and BA have been separately proposed as potential ancestral nucleobases of proto-RNA for additional reasons [13, 14]. Here, we show that MA and BA spontaneously form glycosidic bonds in water with ribose and ribose-5'-phosphate in excellent yields to produce nucleosides and nucleotides. Moreover, their nucleotides were found to assemble with each other into ordered supramolecular structures that might have facilitated the emergence of proto-RNA polymers.

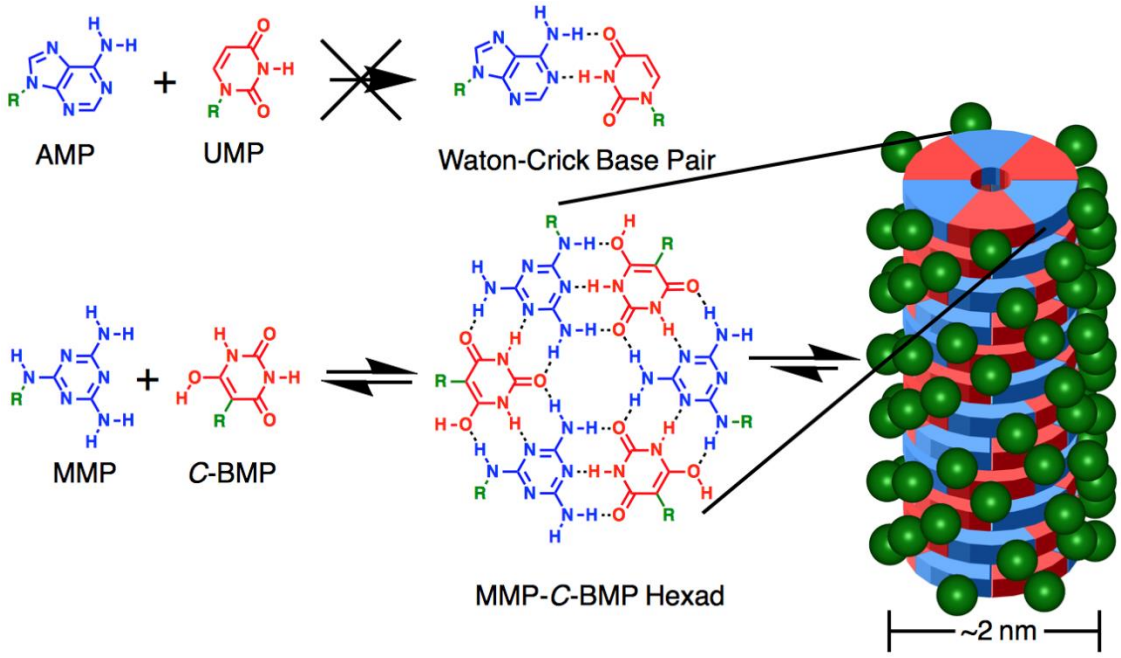


Figure 7.1 Comparison of a canonical base pair to that of the MMP-BMP base pair. Chemical structures of MMP and *C*-BMP and their association into stacked hexads, a structure that has a predicted width that is consistent with structures observed in AFM images. Ribose-5-phosphate is indicated by green R on nucleobase chemical structures and as green sphere on hexad stack. The similarly of AMP to MMP and UMP to *C*-BMP, and the fact that these canonical nucleosides do not assemble in aqueous solution is also illustrated.

7.2 Experimental Procedures

7.2.1 Materials

Melamine and barbituric acid were purchased from Acros Organic, D-ribose-5-phosphate disodium salt and D-ribose were purchased from Sigma Aldrich.

7.2.2 Sample Preparation

Unless otherwise noted, assemblies were formed by combining 50 mM of both melamine and barbituric acid in the form of the parent heterocycles, corresponding nucleotides or both forms at 50 mM in each heterocycle total. Because of the high yielding synthesis of the BMP (between 80-95% yield), the crude reaction mixture was used for spectroscopic analysis without purification. All solutions contained 300 mM in NaCl except when AFM imagine was performed on assemblies containing purified BMP and MMP which contained 1 M NaCl. Solutions were pH adjusted by the addition of NaOH or HCl to pH between 4.5 and 5.0 unless otherwise noted. All solutions gelled immediately upon mixing complementary molecules and adjusting pH, except for MMP+BA assemblies. Experiments evaluating MMP+BA assemblies involved first incubating the assemblies at 5 °C overnight, after this time a hydrogel was present. Anomerization experiments containing MMP were first incubated at 5 °C for 24 hours to enable the anomeric ratio to equilibrate under the conditions tested.

7.2.3 Analytical Techniques

Spectroscopic analysis of the supramolecular assemblies was performed using circular dichroism (CD) and ^{1}H NMR. CD analysis was carried out on a Jasco J-720 CD

spectrometer equipped with a six cell Quantum Northwest peltier temperature controller. Strain free 0.01 mm demountable cells from Starna were used. NMR spectra were collected on a Bruker DRX-500 500 MHz NMR and were the sum of 32 transients. All molecules were D_2O exchanged and lyophilized prior to analysis in D_2O with an internal standard of TSP at 1.11 mM.

AFM imaging was performed with a Nanoscope IIIa (Digital Instruments) in tapping mode, using Si tips (Vistaprobes, 40 N/m). A freshly cleaved mica substrate was pre-activated by incubation with a solution of 20 mM $MgCl_2$, which was then rinsed with water and dried under N_2 (g). A 3 ul sample was deposited on the mica substrate and spread with N_2 gas followed by the addition of 50 uL of cold water to remove excess sample.

7.2.4 Synthesis of 5-Ribofuranosyl-Barbiturate-5′-Monophosphate (BMP)

Barbituric acid (2.5 mmol) and ribose-5-phosphate (2.5 mmol) were dissolved in 5 mL of H_2O and the pH was adjusted to 9.0 with sodium hydroxide. The solution was stirred for 24 hours at 20 °C. This solution was then loaded onto a gravity column containing QAE Sephadex A-25 anion exchange media and eluted with a gradient of NH_4HCO_3 buffer from 50 mM to 0.5 M. The fractions containing product were combined. Maximum yield of BMP was 90% (α-BMP 26%, β-BMP 64%). HRMS (neg. *m*/*z*): $C_9H_{12}N_2O_{10}P^-$ theoretical mass: 339.0235, actual mass: 339.0243.

***α*-BMP**: ^{1}H NMR (500 MHz, D_2O) δ 4.94 (d, 4.2 Hz, H1'); 3.85 (dd, 4.2, X Hz, H2'); 3.93 (dd, 4.1, 3.5 Hz, H3'); 3.70 (m, H4'); 3.59 (m, H5'a); 3.46 (m, H5'b). ^{13}C NMR (126

MHz, D_2O) δ 167.1 (C4/C6); 152.7 (C2); 85.5 (C5); 79.5 (d, 8.2 Hz, C4'); 75.1 (C1'); 74.3 (C2'); 72.7 (C3'); 63.6 (d, 4.4, C5').

***β*-BMP**: ^{1}H NMR (500 MHz, D_2O) δ 4.51 (d, 5.7 Hz, H1'); 4.29 (t, 5.7 Hz, H2'); 3.86 (t, 6.2 Hz, H3'); 3.54 (m, H4'); 3.58 (m, H5'a); 3.46 (m, H5'b). ^{13}C NMR (126 MHz, D_2O) δ 166.4 (C4/C6); 153.1 (C2); 85.7 (C5); 80.8 (d, 7.9 Hz, C4′); 78.6 (C1'); 70.9 (C2'); 70.5 (C3'); 63.8 (d, 4.6 Hz C5').

7.2.5 Synthesis of *N*-Ribofuranosyl-Melamine-5'-Monophosphate (MMP)

Melamine (1 mmol) and ribose-5-phosphate (1 mmol) were dissolved in 5 mL of H_2O and the pH was adjusted to 5.0 with hydrochloric acid. The solution was stirred at 65°C for 24 hours. This solution was then loaded onto a gravity column containing SP Sephadex C-50 cation-exchange media and eluted with NH_4OAc buffer, 50 mM, pH 4.31. The fractions containing product were lyophilized, redissolved in water, and pooled. Maximum yield of **MMP** was 55% (*α*-**MMP** 26.4%, *β*-**MMP** 28.6%). HRMS (neg. m/z): $C_8H_{14}N_6O_7P^-$ theoretical mass: 337.0667, actual mass: 337.0659.

***α*-MMP**: ^{1}H NMR (500 MHz, D_2O) δ 5.46 (d, 4.1 Hz, H1'); 3.97 (dd, 4.7, 1.9 Hz, H3'); 3.94(dd, 4.7, 4.1 H2'); 3.74 (m, H4'); 3.64 (m, H5'a); 3.55 (m, H5'b). ^{13}C NMR (126 MHz, D_2O) δ 164.4 (C2); 164.2 (C3/C4); 81.0 (C1'); 79.5 (d, 8.2 Hz, C4'); 70.0 (C3'); 69.9 (C2'); 63.4 (d, 5.3, C5').

***β*-MMP**: ^{1}H NMR (500 MHz, D_2O) δ 5.25 (d, 6.6, H1'); 3.89 (dd, 5.4, 3.2 Hz, H3'); 3.85 (dd, 6.6, 5.4 Hz, H2'); 3.75 (m, H4'); 3.49 (m, 3.54, H5'a,b). ^{13}C NMR (126 MHz, D_2O) δ 165.1 (C2); 164.9 (C3/C4); 84.2 (C1'); 82.3 (d, 8.0 Hz, C4'); 73.1 (C2'); 70.5 (C3'); 64.2 (d, 4.5 Hz, C5').

7.3 Results and Discussion

7.3.1 Glycosylation of Barbituric Acid and Melamine by Ribose

Glycosylation of BA by one equivalent of ribose (500 mM each) at 20 °C in water (without drying) was found to be almost quantitative, with BA+ribose conjugates exceeding 90%. The products of this reaction include both the α and β anomers of the furanosyl and pyranosyl forms of ribose (Figure 7.2). The pyranosyl form appears to be produced in the greatest yield (see overlapping pair doublets between 4.2 and 4.3 ppm in Figure 7.2C; however sugars corresponding to these assignments were not fully characterized). The finding that the pyranoside is formed preferentially over the furanoside is intriguing since the other two reported pyrimidine glycosylation reactions from our lab (i.e., 2-pyrimidinone and TAP) both formed the furanoside as the major product. Recently, Benner and coworkers demonstrated that 5-aminouracil can be efficiently glycosylated by ribose at the 5C position of the pyrimidine ring (producing the *C*-nucleoside) with the major products being the pyranosides produced at a ratio of 3:1 [15]. A more detailed investigation is needed in order to evaluate why some pyrimidines favor one sugar form over the other.

Reactions between MA and one equivalent of ribose were also productive, forming four MA-ribose conjugates in similar yield with an almost 30% total yield when reactions were carried out at 65 °C (Figure 7.3). Previously, Tor and coworkers reported that *N*-methylated melamine (attached to ribose at the same methylated exocyclic amine position) rapidly hydrolyzed in water before the product could be identified by NMR [13]. The observation that the melamine nucleosides described here were able to be identified even after water-based purification suggests that the stability of the methylated

form of melamine nucleosides is poorer and that melamine would be a more advantageous proto-RNA recognition element.

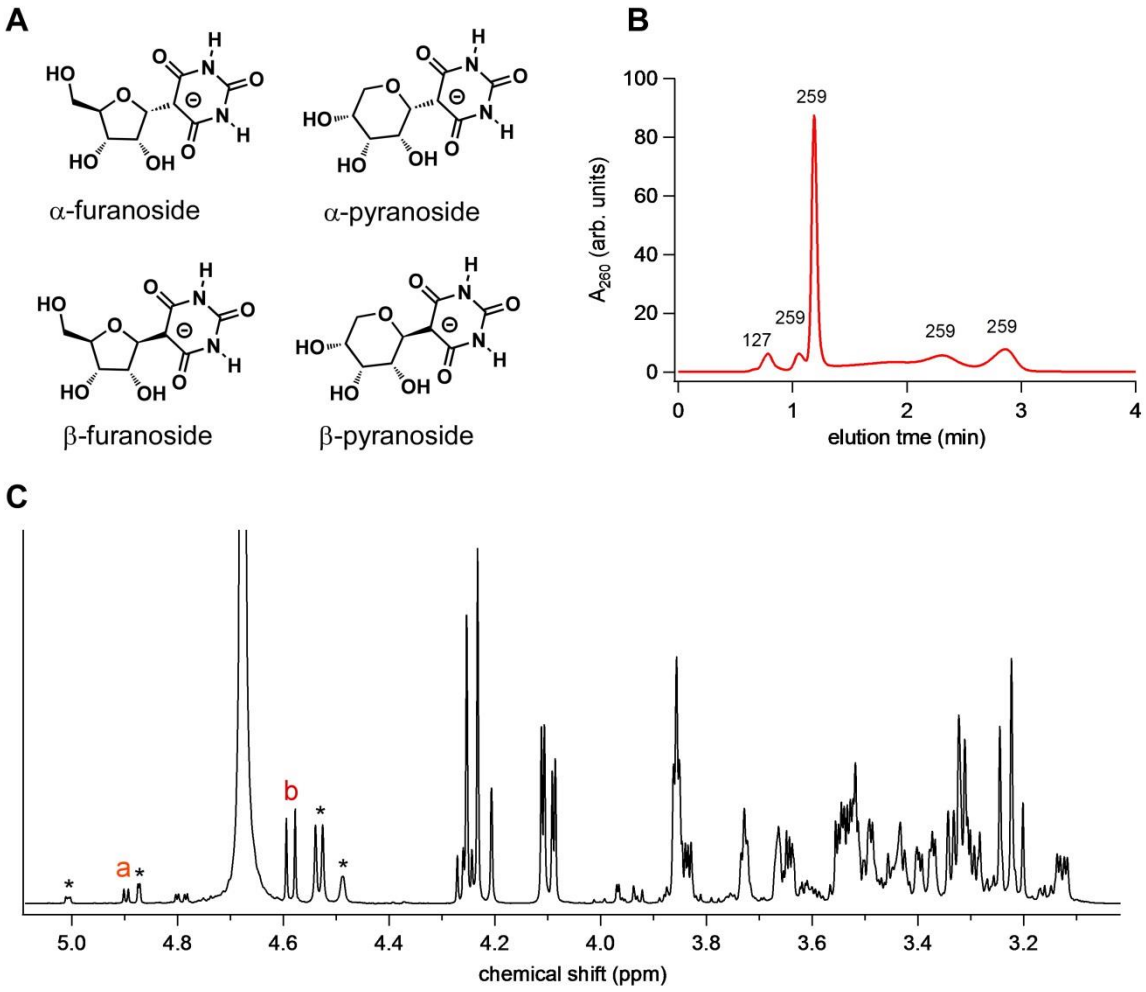


Figure 7.2 ^{1}H NMR and HPLC analysis of the BA+ribose reaction. (A) Chemical structure of BA+ribose conjugates that can be formed by reacting BA with ribose. (B) HPLC chromatograph of samples of BA and ribose after reacting at 20 °C for 24 h in water. Peaks are labeled with m/z values obtained by simultaneous MS (negative mode) and UV absorption monitoring of LC. The m/z values listed correspond to the deprotonated BA heterocycle, 127; and four peaks with the mass of BA with a closed-ring ribose conjugate, 259. (C) ^{1}H NMR of the BA+ribose reaction depicting the nonexchangeable ribose protons. The anomeric protons of the four unconjugated ribose sugar forms are labeled (*) and the α and β BA-ribofuranosides are labeled α and β, respectively.

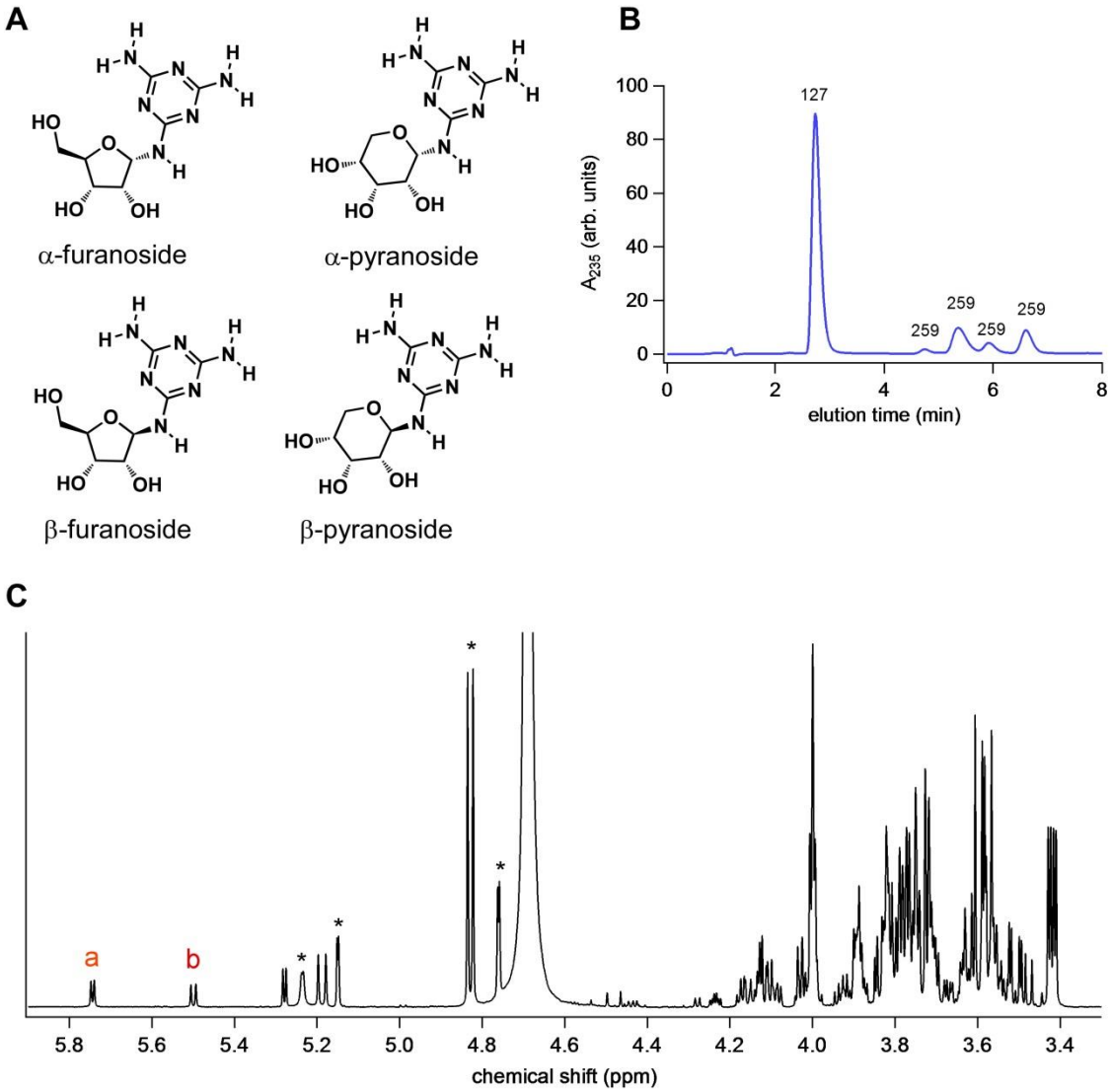


Figure 7.3 ^{1}H NMR and HPLC analysis of the MA+ribose reaction. (A) Chemical structure of MA+ribose conjugates that can be formed by reacting MA with ribose. (B) HPLC chromatograph of samples of MA and ribose after reacting at 65 °C for 24 h in water. Peaks are labeled with m/z values obtained by simultaneous MS (positive mode) and UV absorption monitoring of LC. The m/z values listed correspond to the MA heterocycle, 127; and four peaks with the mass of MA with a closed-ring ribose conjugate, 259. (C) ^{1}H NMR of the MA+ribose reaction depicting the nonexchangeable ribose protons. The anomeric protons of the four unconjugated ribose sugar forms are labeled (*) and the α and β MA-ribofuranosides are labeled α and β, respectively.

7.3.2 Glycosylation of Barbituric Acid and Melamine by Ribose-5-Phosphate

Previous reports that phosphorylated sugars can be produced in model prebiotic reactions motivated us to explore the potential for BA and MA to be glycosylated by phosphorylated ribose to provide an abiotic route to nucleotides [16]. When ribose-5-phosphate (R5P) was evaluated in place of ribose, similar yields were obtained for BA glycosylation, i.e., 82% (Figure 7.4A). Surprisingly, MA glycosylation yields doubled, to 55%, when R5P was used in place of ribose in reactions with MA (Figure 7.4B). The higher yields form melamine glycosylation by R5P compared to ribose may be due to charge pairing between the positively charged melamine (pK_a 4, reaction performed at pH 5) and monoanionic phosphate of R5P. The number of products observed in the R5P reactions are less than that observed in the ribose reactions, as R5P excludes the formation of the pyranosides due to R5P favoring the furanose form of ribose. These high yields are extraordinary among model prebiotic reactions, especially considering that none of the four canonical nucleobases form nucleosides when heated with ribose in water. On the contrary, the canonical nucleosides and nucleotides are thermodynamically unfavored (but kinetically stable) in water [17].

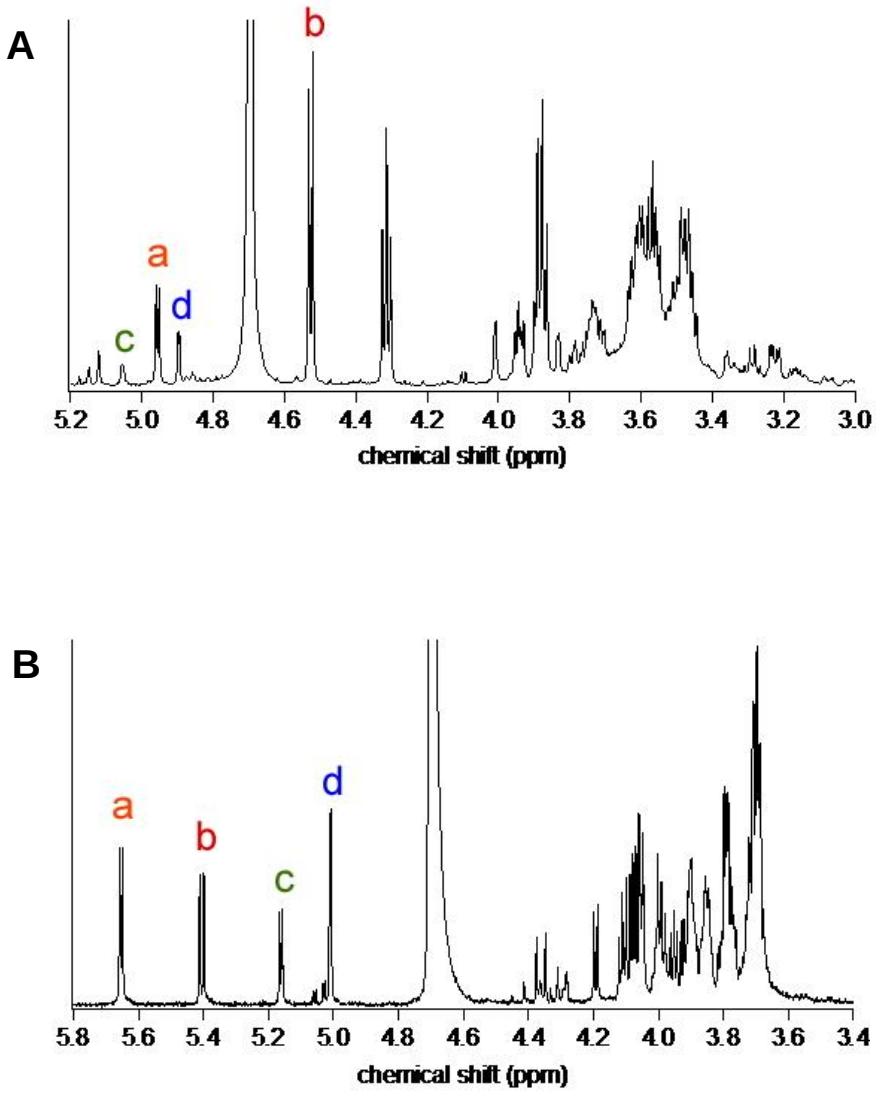


Figure 7.4 ^{1}H NMR analysis of unpurified nucleosidation reactions between ribose-5-phosphate and BA or MA. (A) ^{1}H-NMR of BA+R5P crude reaction mixture. The anomeric protons are indicated for a, α-BMP; b, β-BMP; c, α-R5P and d, β-R5P. (B) NMR of BA+ribose-5-phosphate crude reaction mixture. The anomeric protons are indicated for a, α-MMP; b, β-MMP; c, α-R5P and d, β-R5P.

7.3.3 Characterization of the Barbituric Acid Nucleotides

Products from the BA reactions with R5P were isolated by anion exchange column chromatography. Analysis by 2D NMR spectroscopy (HMBC, HSQC and COSY) was used to confirm the identity of the BA+R5P conjugates (Figure 7.5). The anomeric protons were found to be up field to where *N*-nucleoside anomeric protons are generally, this observation along with through bond correlations (HMBC) between the anomeric protons and the C5 carbon of the aromatic pyrimidine ring (approx. 85 ppm) suggest the formation of *C*-nucleosides, with a C-C glycosidic bond between ribose and BA (BMP). The *C*-nucleoside was expected due to the well-known nucleophilicity of the C5 carbon of barbiturate. In an earlier study, glucose, another biological aldose, was shown to react with BA in high yield (approx. 80%) when refluxed in water forming the *C*-glycosyl product [18]. 1D ROE spectroscopy was used to confirm that the β-anomer is preferentially formed in a 1:2 ratio to the α-anomer (Figure 7.6). Magnetization transfer from the major anomeric proton to a 4' proton with TOCSY transfer (through bond magnetization transfer) from the H1' to the H2' and H3' confirmed the β-anomer assignment. While the α-anomer shows magnetization transfer from the H1′ to the H2′ and H3′. Intermediate products (specifically, an aminated side-product of the Knovenagel intermediate) suggest that the BA nucleotide-formation reaction proceeds through a Knovenagel condensation (Figure 7.7). We note that β-BMP shares a close structural relationship with the *C*-nucleotide pseudouridine, the most common posttranscriptionally modified nucleotide in RNA [19].

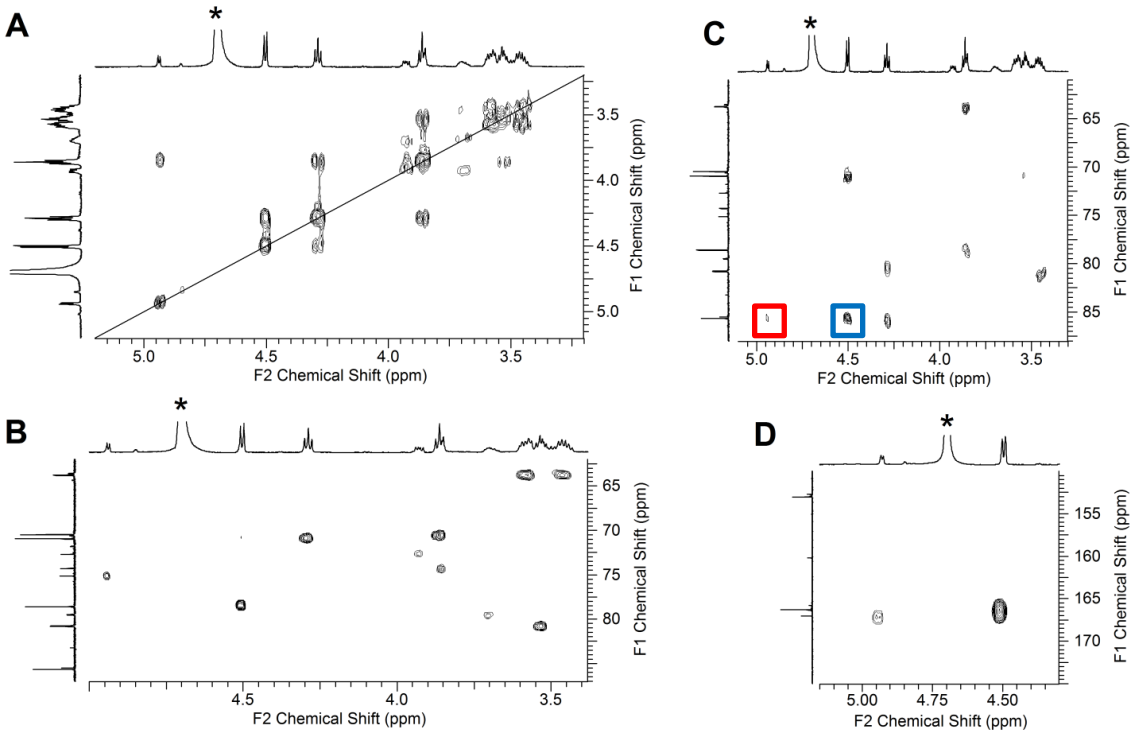


Figure 7.5 2D NMR of both anomers of BMP after purification. (A) COSY, (B) HSQC and (C,D) HMBC spectra used to confirm proton and carbon assignments of α-BMP and β-BMP. As with other pyrimidine *C*-nucleosides the anomeric proton resonances are generally up field (ca. 4.5-5 ppm) [10, 20, 21] in relation to *N*-nucleosides. Panel C shows correlation between the anomeric protons and C5 carbon of barbituric acid, boxed in red α-BMP and blue for β-BMP. Panel D shows correlation between the anomeric protons and the C4 and C6 position of barbituric acid.

* indicates HOD.

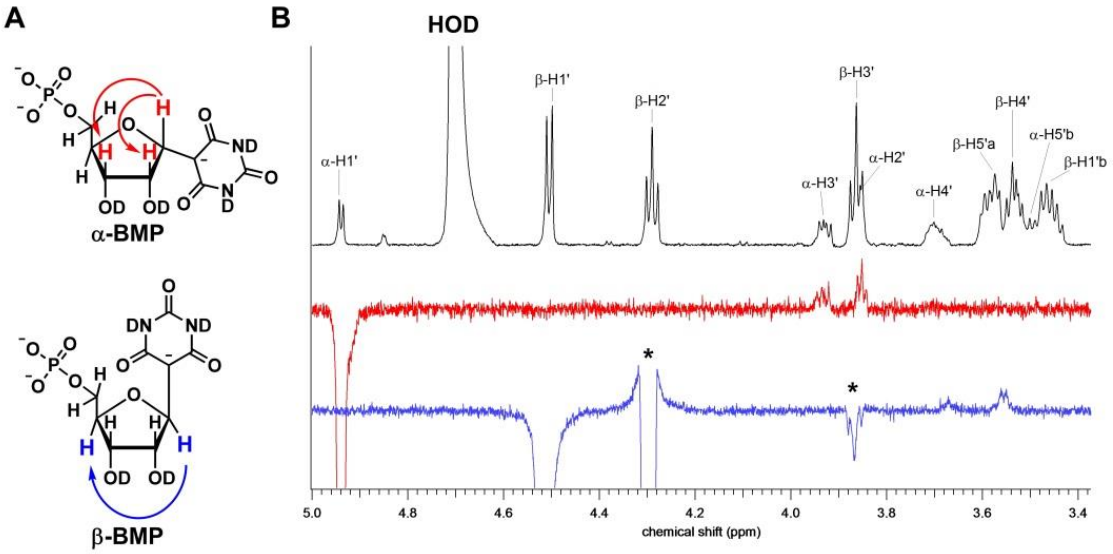


Figure 7.6 1D ROE analysis of purified BMP. (A) Chemical structure of both anomeric forms of BMP. Arrows indicate through space proton-proton magnetization transfer as indicated by ROE analysis. (B) ^{1}H NMR and ROE spectra of non-exchangeable protons (ribose) of both anomers of BMP. Resonance assignments determined by 2D NMR analysis are indicated above the ^{1}H NMR spectrum. Red trace (middle) corresponds to irradiation of the anomeric proton (H1') of α-BMP resulting in through space magnetization transfer to H2' and H3'. Blue trace (bottom) shows irradiation of the anomeric proton (H1') of β-BMP resulting in through-space magnetization transfer to H4'. Through-bond magnetization transfer (TOCSY) is indicated by (*).

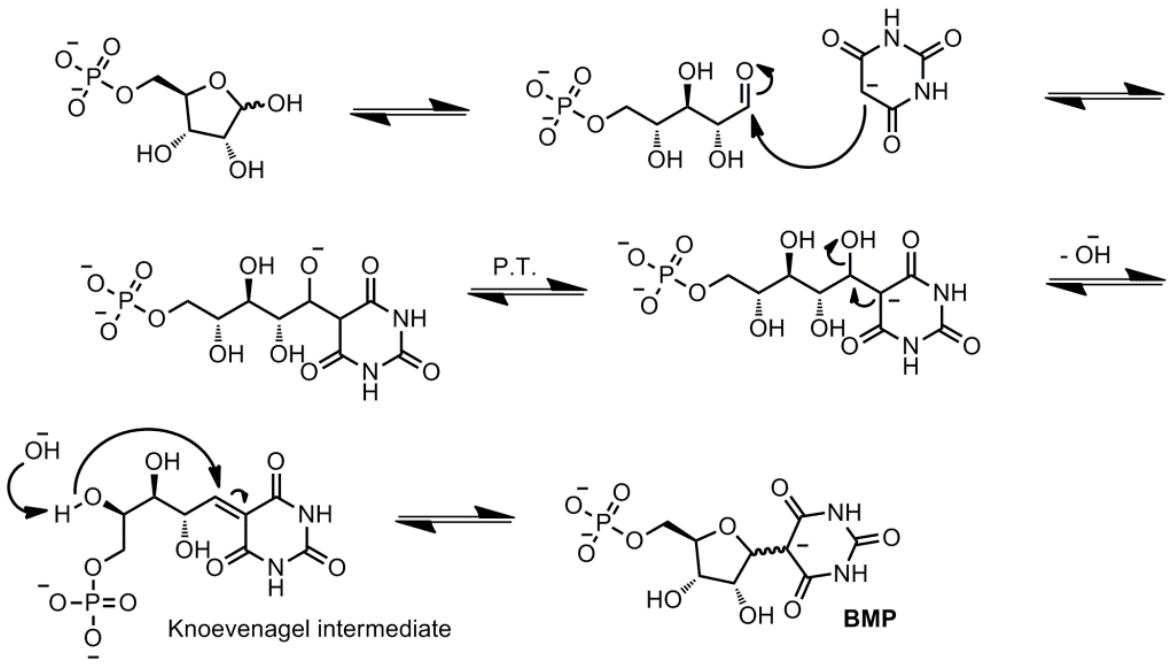


Figure 7.7 Proposed mechanism for BMP formation from ribose-5-phosphate and barbituric acid. A Knoevenagel condensation results from attack of the nucleophilic C5 of barbituric acid on the aldehyde of ribose-5-phosphate when the sugar is in its open form. Next, the C(5′)-OH group attacks the enone causing ring closure and formation of both anomers of the *C*-nucleoside. Note that the charge state of the molecules reflects the reaction at pH 9. Mechanism proposed by D. Fialho.

7.3.4 Characterization of the Melamine Nucleotides

Products from the MA reactions with R5P were isolated by cation exchange column chromatography. NMR analysis of the MA+R5P conjugates was also performed using HMBC, HSQC, COSY and ROEs (Figures 7.8 and 7.9). As predicted, glycosylation of MA was found to occur at the N2 position forming the exocyclic amino nucleotides of melamine (MMP). Both the α and β anomers were found to be in approximately equal ratio at equilibrium at 5 °C (0.45 to 0.55, α and β, respectively). The reaction mechanism, see Figure 7.10, shows that the reaction proceeds through a reversible Schiff base intermediate followed by ring closer resulting in formation of both furanosyl anomers (Figure 7.10). Similar anomeric ratios near unity were also observed with exocyclic amino nucleosides formed between R5P and TAP (not shown) which stands in contrast to both the *C*-nucleosides of TAP and BA favoring the β-anomer (likely due to greater steric hinderence caused by linking the aromatic ring directly to the sugar). The melamine nucleosides were found to be highly stable in water with a half-life of approx. 200 days at 5 °C and pH 5 (Figure 7.11). This was surprising in light of a report by Tor and coworkers showing that exocyclic amino triazine nucleosides of N2,N4,N6-trimethyl-melamine hydrolyses completely in water within 15 minutes through a Schiff base intermediate that also results in anomerization even at higher pH (from pH 7-10) [13]. Since the degradation study was carried out at pH 5 the stability of the nucleotide may be much greater at increased pH.

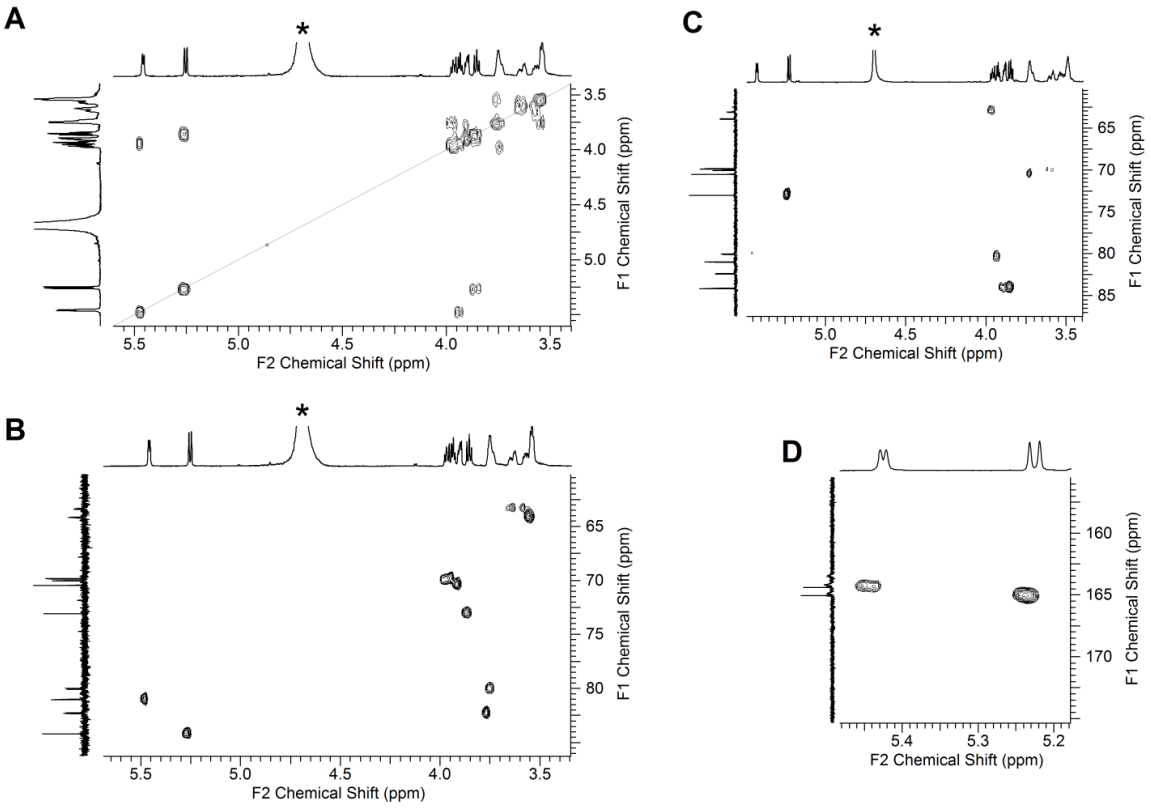


Figure 7.8 2D NMR of both anomers of MMP after purification. (A) COSY, (B) HSQC and (C,D) HMBC spectra used to confirm proton and carbon assignments of α-MMP and β-MMP.
* indicates HOD.

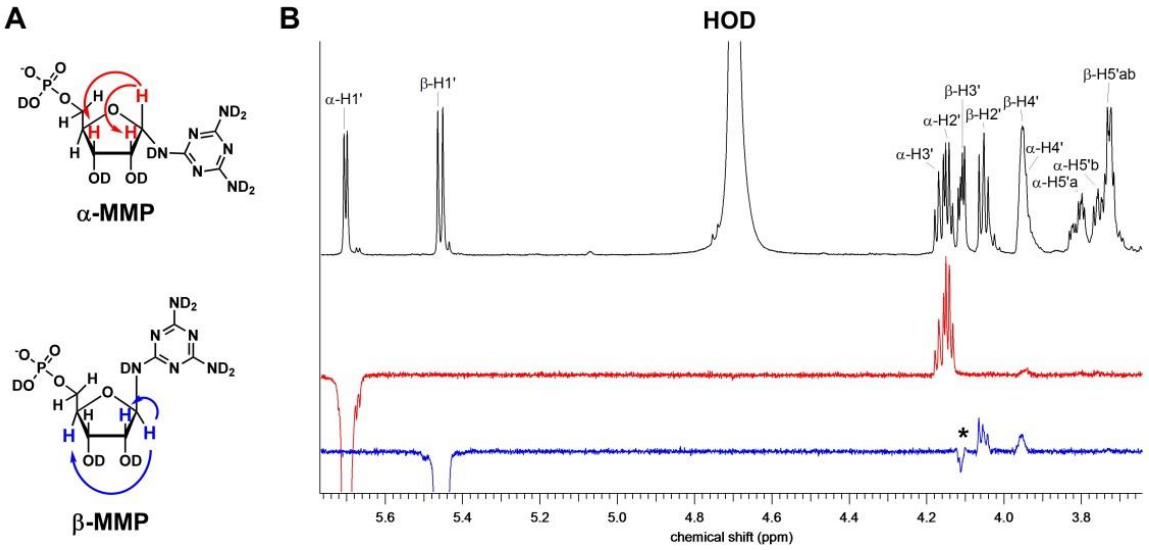


Figure 7.9 1D ROE analysis of purified MMP. (A) Chemical structure of both anomers of MMP. Arrows indicate through space proton-proton magnetization transfer as indicated by ROE analysis. (B) ^{1}H NMR and ROE spectra of non-exchangeable protons of MMP. Resonance assignments determined by 2D NMR analysis are indicated above the ^{1}H NMR spectrum. Red trace (middle) corresponds to irradiation of the anomeric proton (H1') of α-MMP resulting in through space magnetization transfer to H2' and H3'. Blue trace (bottom) shows irradiation of the anomeric proton (H1') of β-BMP resulting in through-space magnetization transfer to H2' and H4'. Through-bond magnetization transfer is indicated by (*).

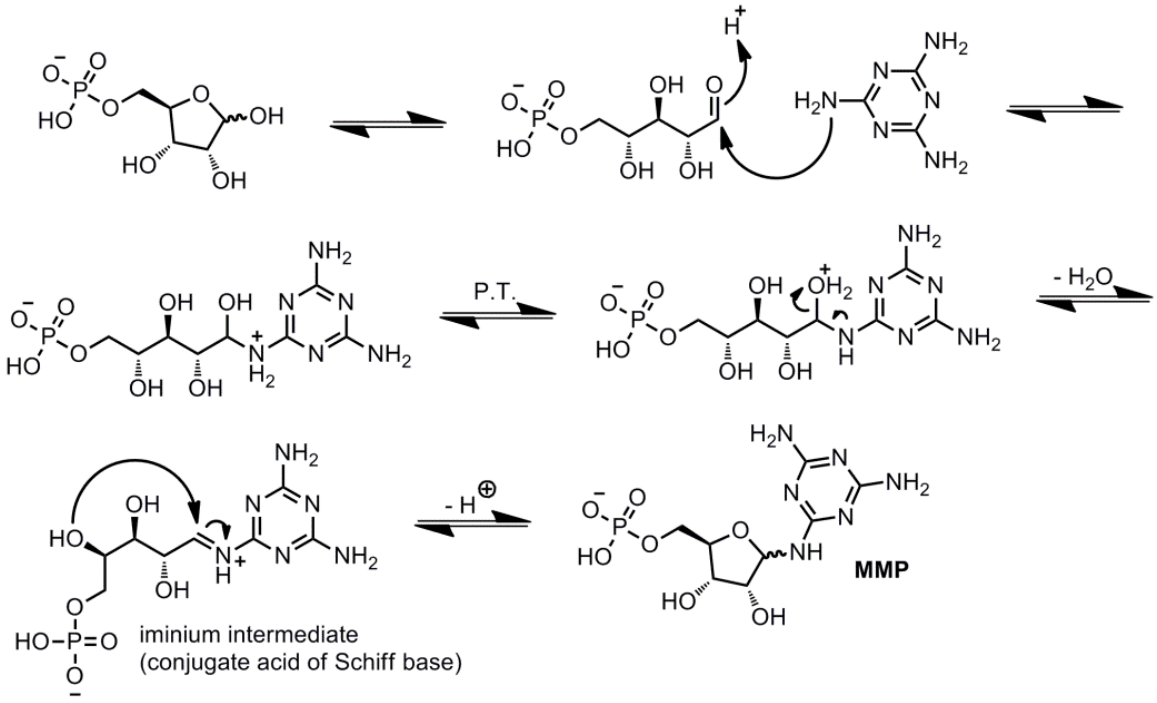


Figure 7.10 Proposed mechanism for MMP formation from ribose-5-phosphate and MA. A Schiff-base is formed from attack of one of melamine's exocyclic amines onto the aldehyde of ribose-5-phosphate followed by elimination of water. The iminium intermediate is attacked by the C(5')-OH leading to ring closure and formation of both anomers of the exocyclic amino nucleotides. Note that the charge state of the molecules reflects the reaction at pH 5. Mechanism proposed by D. Fialho.

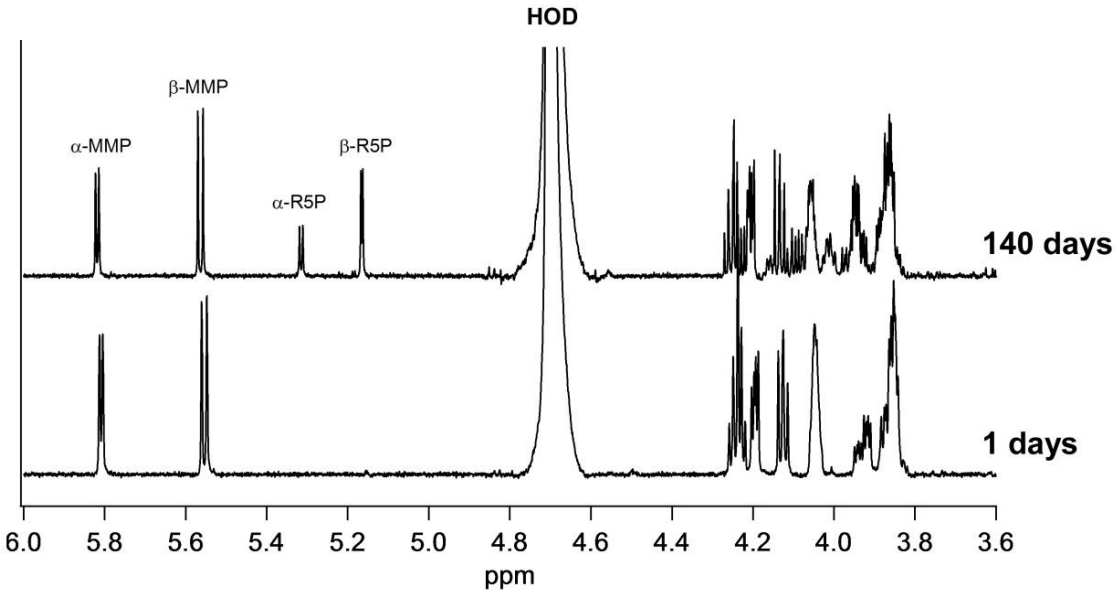


Figure 7.11 MMP hydrolysis over 140 days in water at 5 °C. ^{1}H-NMR analysis of a 5 mM MMP solution after 1 and 140 days (anomeric protons of both anomers of MMP and R5P are labeled). The solution was originally adjusted to pH5 and contained 300 mM NaCl was stored at 5 °C during the course of the experiment.

7.3.5 Supramolecular Polymers Form From Nucleotide Co-Assembly

We next tested the ability of the MMP and BMP to direct their own selection and organization in solution, a property that could have facilitated the co-localization of the nucleotides on the prebiotic Earth and perhaps facilitated their polymerization. When purified or crude reaction mixtures of MMP and BMP were combined, a shear-thinning hydrogel formed indicative of the formation of long noncovalently cross-linked assemblies (Figure 7.12). These assemblies were not present in the reaction mixtures on their own or within solutions of purified MMP and BMP, indicating that the gels were forming due to co-assembly of both heterocycles and their derivatives. Circular dichroism (CD) analysis of similar mixtures containing equal amounts of either MMP, in the crude reaction or isolated, with BMP or BA, or MA with BMP, indicated the formation of temperature responsive helical supramolecular assemblies (Figures 7.13-7.16). For each solution, visualization of the assemblies by atomic force microscope (AFM) revealed the incorporation of thousands of base paired nucleotides into long supramolecular polymers with diameters of approx. 2 nm, fully consistent with the presence of stacked hydrogen-bonded hexad-based "rosette" structures described previously for assemblies between modified TAP and Cy heterocycles (Figures 7.17-7.20) [10]. The supramolecular polymers formed from purified MMP and purified BMP were found to be the markedly shorter than the other systems, this likely results from frustrated growth of the more charged assemblies (containing six charges per rosette as opposed to three charges as is the case for assemblies of MMP+BA and BMP+MA) [22].

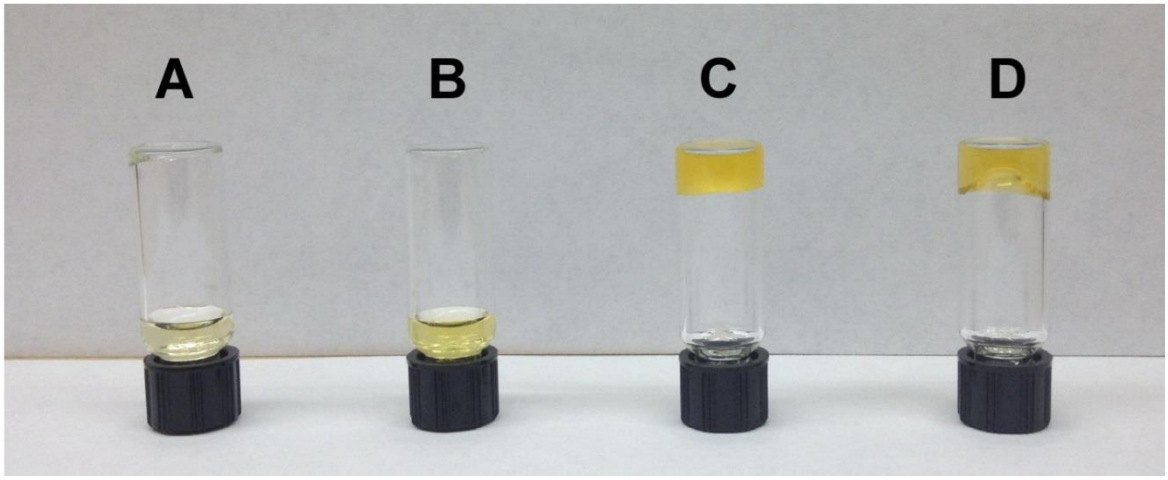


Figure 7.12 Inverted bottle test demonstrating hydrogel formation of solutions that contain both MMP and BMP. (A) MMP reaction solution at 50 mM (B) BMP reaction solution at 50 mM (C) MMP and BMP reaction solutions combined at 50 mM each (note opaque quality likely results from inclusions of precipitates formed by MA and BA heterocycles) and (D) purified MMP and BMP reaction solution combined at 50 mM each. All solutions contained 300 mM NaCl and were pH adjusted to 4.5.

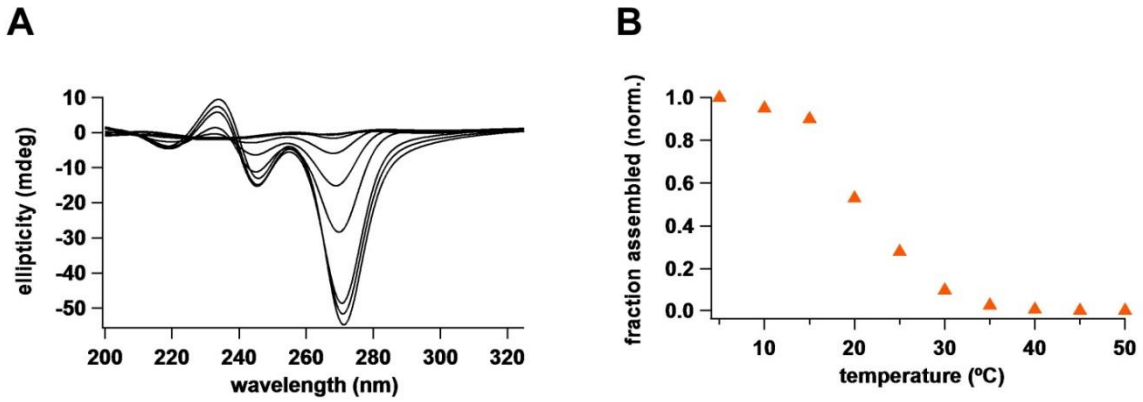


Figure 7.13 CD analysis of a solution containing BMP and MA. (A) CD spectra of a solution (50 mM in each molecule; 300 mM NaCl; pH 7), ranging in temperatures from 5 to 50 °C. (B) Plot derived from spectra in panel A of the normalized fraction of molecule assembles as a function of temperature. Unlike the other solutions tested the pH was adjusted to 7.3 instead of 4.5 due to slight precipitation at lower pH.

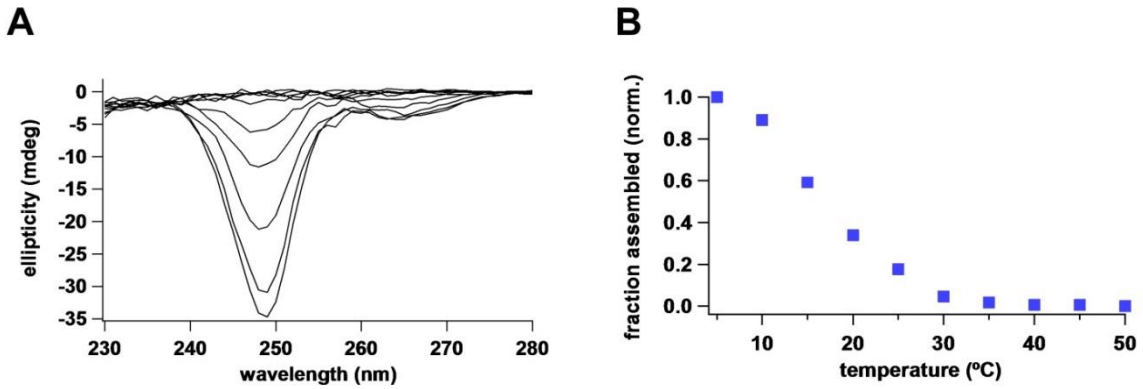


Figure 7.14 CD analysis of a solution containing purified MMP and BA. (A) CD spectra of a solution (50 mM in each molecule; 300 mM NaCl; pH 7), ranging in temperatures between 5 to 50 °C. (B) Plot derived from spectra in panel A of the normalized fraction of molecule assembles as a function of temperature.

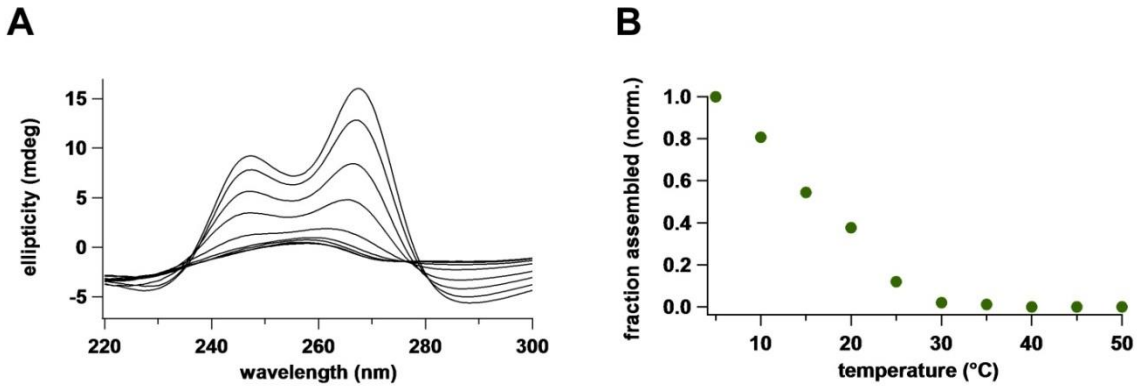


Figure 7.15 CD analysis of a solution containing purified MMP and BMP. (A) CD spectra of a solution (50 mM in each molecule; 300 mM NaCl; pH 7), ranging in temperatures from 5 to 50 °C. (B) Plot derived from spectra in panel A of the normalized fraction of molecule assembles as a function of temperature.

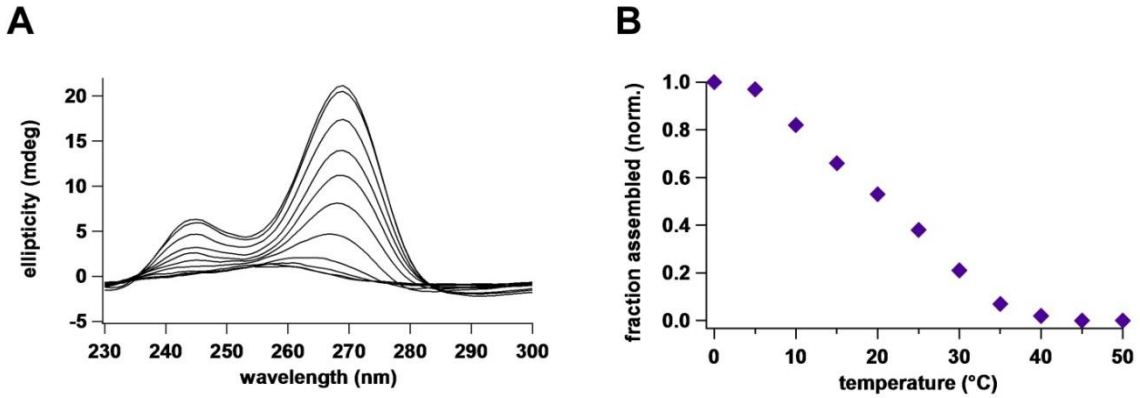


Figure 7.16 CD analysis of a solution containing crude MMP and BMP. (A) CD spectra of a solution (50 mM in each molecule; 300 mM NaCl; pH 7), ranging in temperatures between 5 to 50 °C. (B) Plot derived from spectra in panel A of the normalized fraction of molecule assembles as a function of temperature. These solutions were passed through a 0.2 µm filter prior to CD analysis due to the presence of some precipitate after combining the two molecules (this is likely due to formation of insoluble melamine-barbituric acid aggregates).

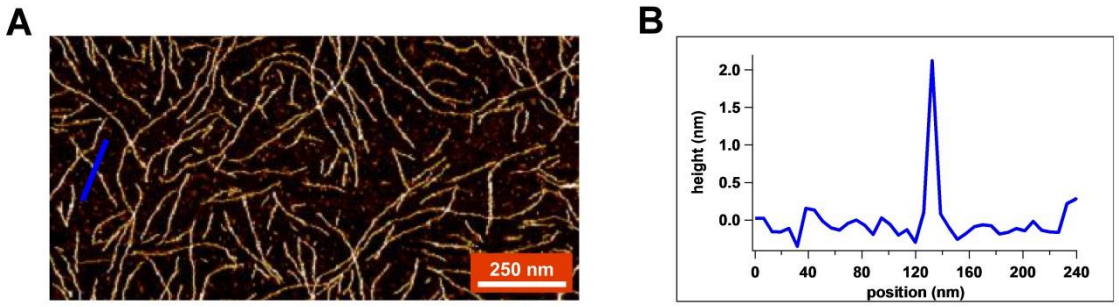


Figure 7.17 MMP and BA assemble into supramolecular polymers. (A) AFM topographic image of assemblies formed by MMP and BA. (B) Height measurements across a linear supramolecular assembly indicated by a blue line in A. Solution contained 50 mM of both MMP and BA and 300 mM NaCl and incubated on ice before depositing on mica.

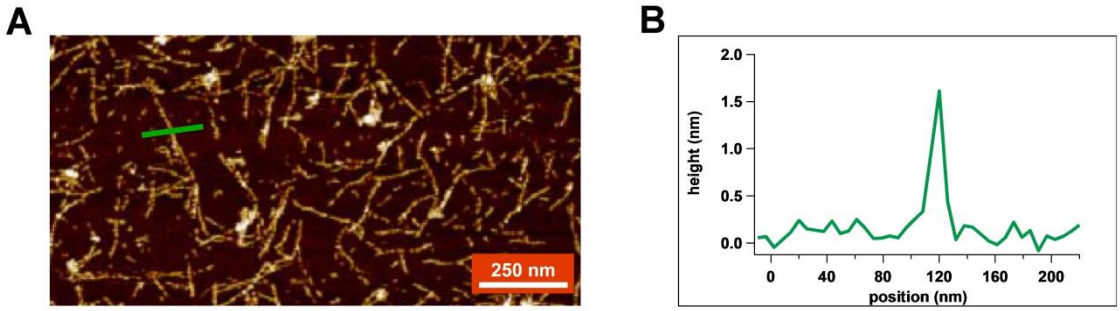


Figure 7.18 BMP and MA assemble into supramolecular polymers. (A) AFM topographic image of assemblies formed by BMP and MA. (B) Height measurements across a linear supramolecular assembly indicated by a green line in A. Solution contained 50 mM of both BMP and MA and 300 mM NaCl was incubated on ice before depositing on mica.

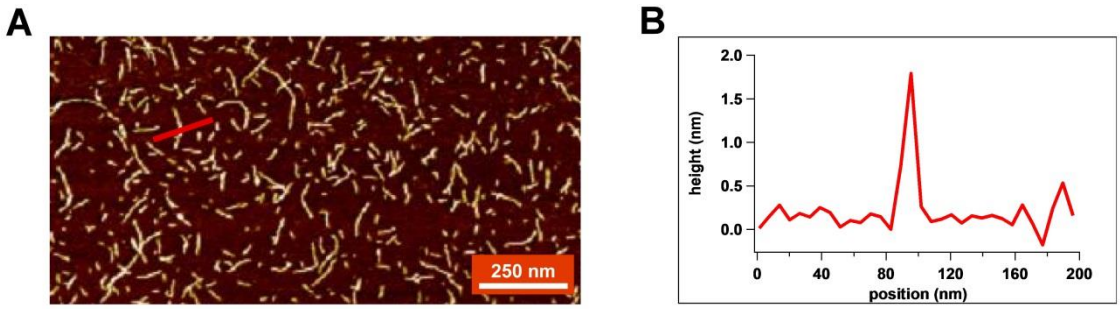


Figure 7.19 Assemblies formed in solutions containing purified MMP and BA. (A) AFM topographic image of assemblies formed by MMP and BA. (B) Height measurements across a linear supramolecular assembly indicated by a red line in A. Solution contained 50 mM of both purified MMP and purified BMP and 1 M NaCl was incubated on ice before depositing on mica. Higher salt was used to enhance assembly between rosettes with a net charge of 6 (compared to the other assembly systems where the rosette net-charge is under 6).

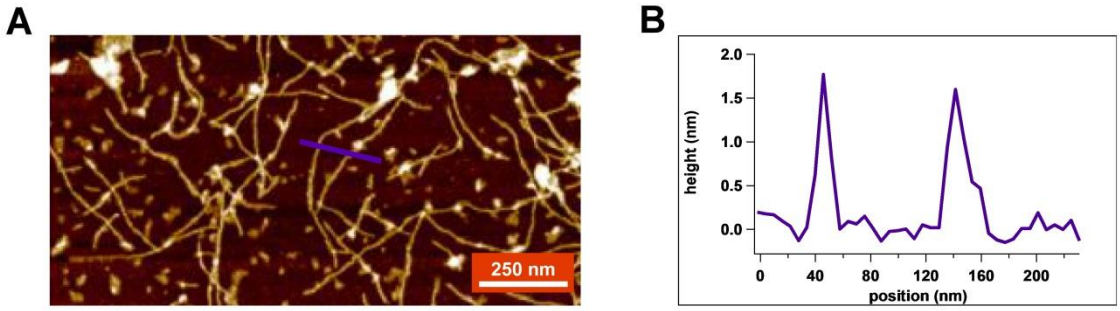


Figure 7.20 Assemblies formed in solutions containing crude MMP and BA. (A) AFM topographic image of assemblies formed by MMP and BA. (B) Height measurements across a linear supramolecular assembly indicated by a red line in A. Solution contained 50 mM of both crude MMP and crude BMP and 300 mM NaCl was incubated on ice before depositing on mica.

7.3.6 β-MMP is Enriched as a Result of Preferential Assembly with Barbituric Acid

We next investigated MMP and BMP assemblies by ^{1}H-NMR spectroscopy. The free nucleotide monomers are in equilibrium with micron-length supramolecular polymers (which experience resonance line broadening that render them invisible to solution state ^{1}H NMR spectroscopy) [23]. Thus, quantitation of free mononucleotides in solution can be used to determine the percent of free and assembled nucleotide [24]. VT-NMR analysis of solutions containing BMP+MA or MMP+BA, 50 mM in nucleotide and heterocycle, showed temperature dependent assembly of nucleotides from 5 to 40°C (Figure 7.21 and 7.22). Preferential incorporation of the β-anomer of MMP into assemblies was observed at all temperatures where the assemblies are present, with a 2-fold preference for β-MMP incorporation at 5 °C (Figure 7.22). As mentioned above, MMP is produced as a mixture of α and β anomers (at equilibrium, 45% α and 55% β). However, when a solution containing 50 mM MMP was incubated with BA at 5°C we observed the conversion of α-MMP to β-MMP after one day, and this conversion reached a maximum after 8 days to 63% of β-MMP (Figure 7.22), a process that was not observed when BA was omitted or when monomer was below the minimal assembly concentration. To promote nucleotide assembly, the ratio of BA to MMP was tested at 3:1 leading to an enhancement of β-MMP, 77% after seven days (Figure 7.23). These observations indicate that MMP is in dynamic exchange between its α and β anomers, and that assemblies formed with BA preferentially select/stabilize the β-form of MMP. Reactions between BA and MMP were observed during these experiments, particularly at 3:1 BA to MMP (see next section). After a number of days this side reaction leads to loss of assembly and a shift in anomeric ratio that approaches that of MMP at equilibrium in solution alone.

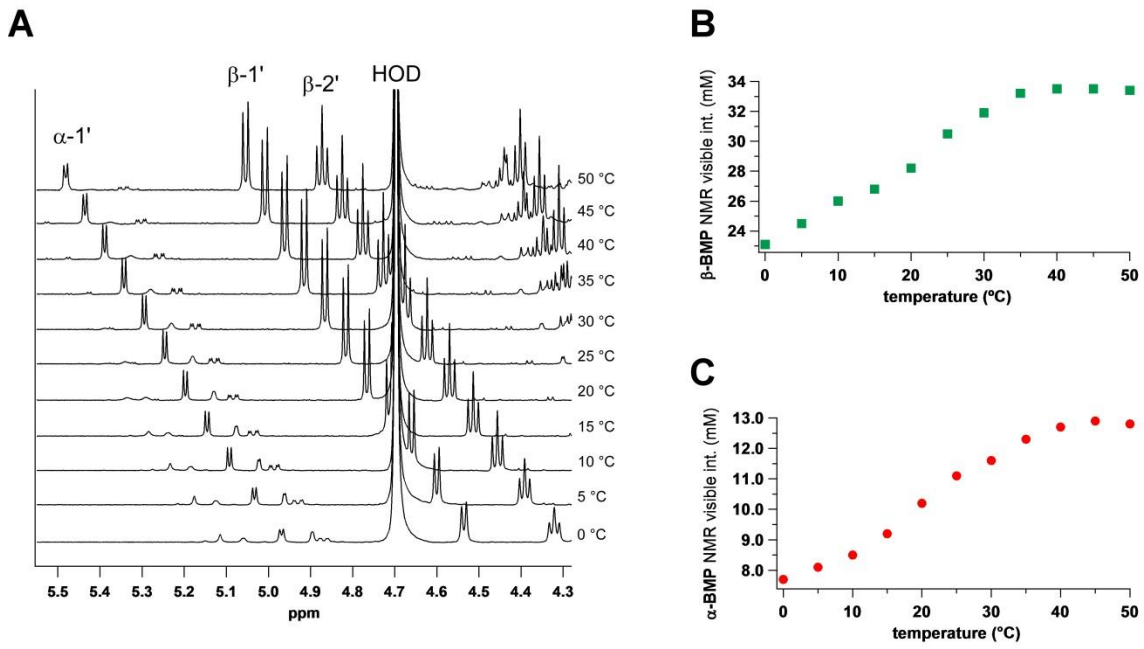


Figure 7.21 VT-NMR analysis of a solution containing BMP and MA. (A) NMR spectra from 0 to 50 °C. (B) NMR visible intensity of *β*- BMP in a solution at various temperatures. (C) NMR visible intensity of *α*-BMP in a solution at various temperatures.

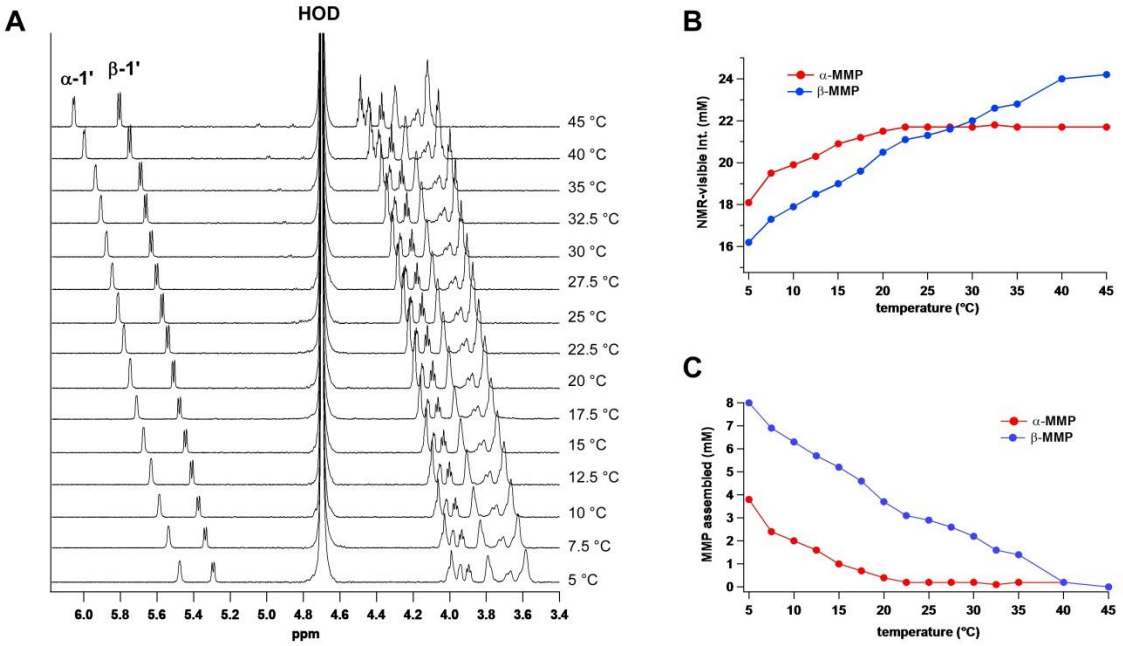


Figure 7.22 VT-NMR analysis of a solution containing MMP and BA. (A) Spectra of a solution containing MMP and BA at temperatures between 5 and 45 °C. (B) Plot of NMR visible resonance intensity on unassembled α and β-MMP as a function of temperature. (C) Amount of each anomer assembled at various temperatures as determined by NMR. Temperatures above 45 °C were not evaluated because of an increase in anomerization rate at higher temperatures. Solutions contained 50 mM MMP and BA as well as 300 mM NaCl and were adjusted to pH 5.

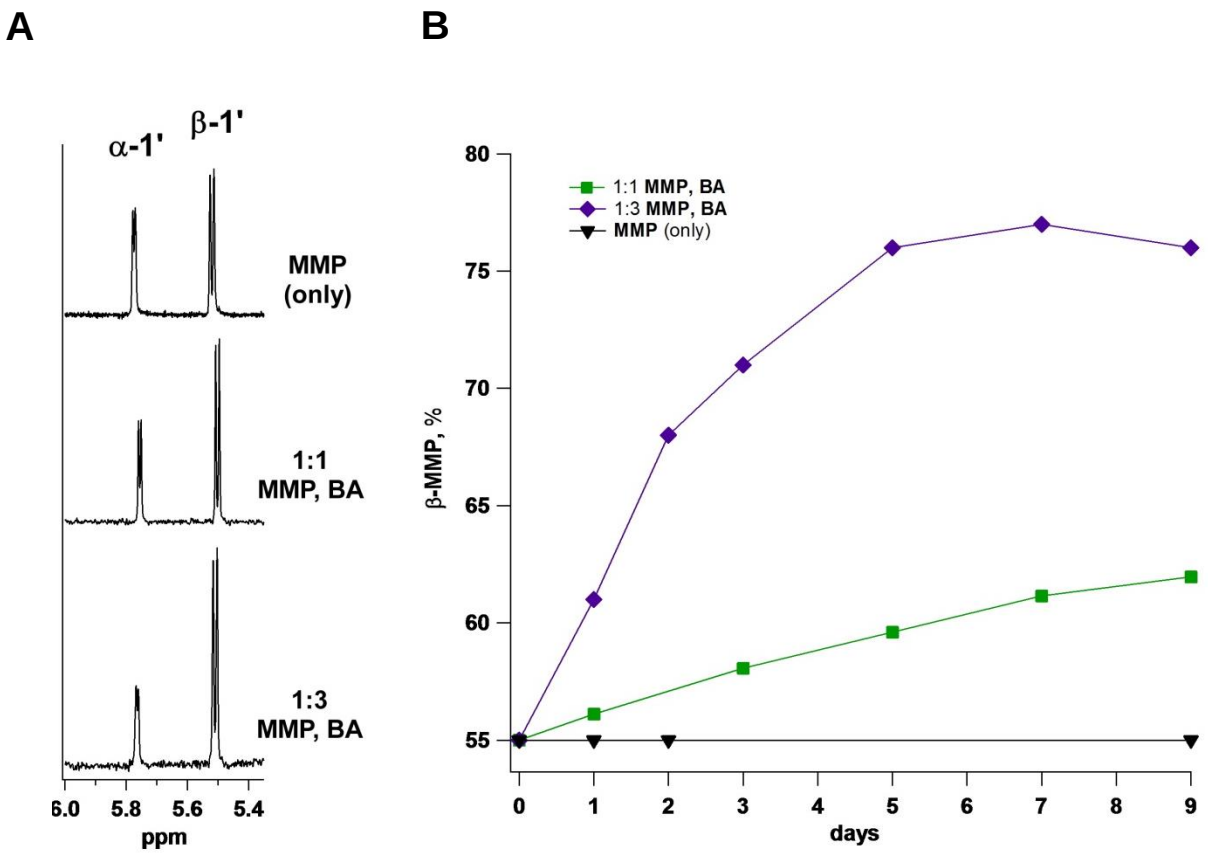


Figure 7.23 ^{1}H NMR analysis of solutions containing MMP and BA over nine days**.** (A) Spectra of solutions containing MMP and various concentrations of BA after nine days storing at 5 °C. (B) Plot showing the percent of *β*-MMP present in solution as a function of time in samples containing both MMP and various concentrations of BA. Note that the bottom spectrum in (A) of 1 BA to 3 MMP has been multiplied by 3.7 in order to maintain the same integrated intensity seen for the top two spectra. This was done because of loss of MMP due to trans-nucleosidation (see next section for more details).

7.3.7 MMP to BMP Trans-Nucleosidation

Evaluation of the supramolecular assembly promoted β-MMP enrichment experiments by ^{1}H-NMR indicated that BA was exchanging with melamine on MMP in solution to form both anomers of BMP (Figure 7.24). This reaction was observed in both the experiment evaluating MMP and BA at equal concentrations (50 mM each) as well as for the experiment involving a solution containing the molecules at a 1:3 ratio (25 mM MMP and 75 mM BA). For solutions containing equivalent concentrations of MMP and BA, BMP represented 13% of the total nucleotide present in solution after incubation at 5 °C for 7 days. The reaction was enhanced in the solution containing three equivalents of BA, constituting 38% of the total nucleotide after incubation for 6 days. Because the hydrolysis of MMP to MA and R5P is slow compared to BMP formation (Figure 7.11), it is likely that BMP is formed by substitution of MA on the linear form of the MA+R5P conjugate (an intermediate of MMP anomerization). Trans-nucleosidation may be responsible for precipitation of solutions containing MMP+BA at 20 °C after 24 hours as higher temperatures would be expected to promote this reaction.

For experiments containing MMP and BA at 1:3 ratio, new resonances were observed in solution after two days that increased in intensity over time. As can be seen in Figure 7.24B, a new resonance located between the anomeric proton signals of BMP appears after the initial formation of BMP and eventually increases to a signal intensity that is larger than the sum of the intensities of both BMP anomers. LCMS analysis of these solutions indicated the presence of double BA addition to R5P (BA_2-R5P). As BMP (i.e., single BA to R5P addition) is an intermediate in the formation of BA_2-R5P, the reaction kinetics observed during NMR analysis also supports this assignment.

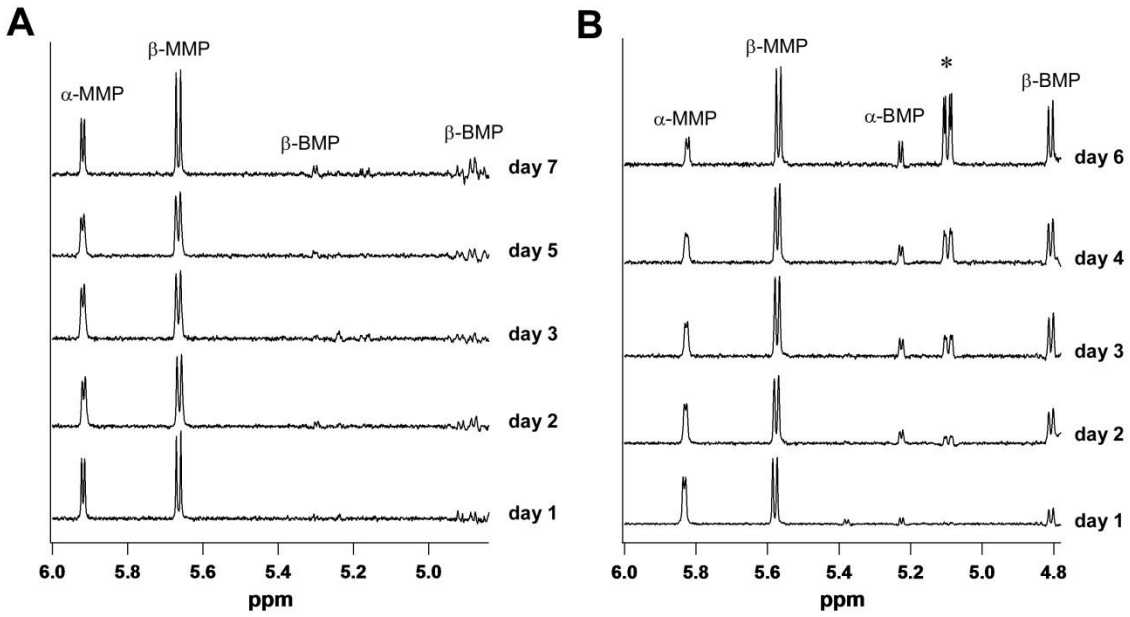


Figure 7.24 NMR analysis of MMP to BMP trans-nucleosidation. (A) Spectra showing the anomeric proton region acquired from a solution originally containing MMP and BA at 50 mM each after 1 to 7 days. (B) Spectra of a solution originally containing 25 mM MMP and 75 mM BA after 1 to 6 days. All solutions contained 300 mM NaCl and were adjusted to pH 5. * indicates proposed BA_2(R5P) assignment.

7.3.8 MMP Assembles with Cyanuric Acid

Melamine and cyanuric acid (Cy) rosette assembly in organic solution has been investigated for 25 years, however, this historically significant assembly motif has not been reported to form in water [12]. Upon mixing Cy and MMP at 50 mM each in water at pH 6 in water a strong hydrogel immediately formed. This pH was chosen as it is at the midpoint between the pK_as of melamine and Cy (pK_a of 5 and 7, respectively). CD analysis of gelled solutions at low temperatures (e.g., 5 °C) showed a very large and negative CD band with a maximum at 225 nm that was lost upon increasing the temperature above 40 °C (Figure 7.25A and 7.25B). VT-NMR revealed that the MAC of both MMP anomers with Cy were about 10-fold lower at 5 °C than with BA and that β-MMP was again preferentially incorporated into the assemblies (Figure 7.25C). These observations prompted us to evaluate the effect of MMP-Cy assemblies on the anomeric ratio of MMP over time. These experiments were carried out at 20 °C as the assemblies were stable at this temperature (which is not the case for MMP-BA assemblies) and a greater ratio of β-MMP to α-MMP is assembled compared to 5 °C. After nine days β-MMP constituted 80% of the total MMP in solution (Figure 7.25D). The greater abundance of β-MMP in solutions when Cy was substituted for BA may result from the enhanced thermal stability of MMP-Cy assemblies (i.e, lower MAC) and larger ratio of β-MMP compared to α-MMP assembled at the temperatures chosen for the different experiments.

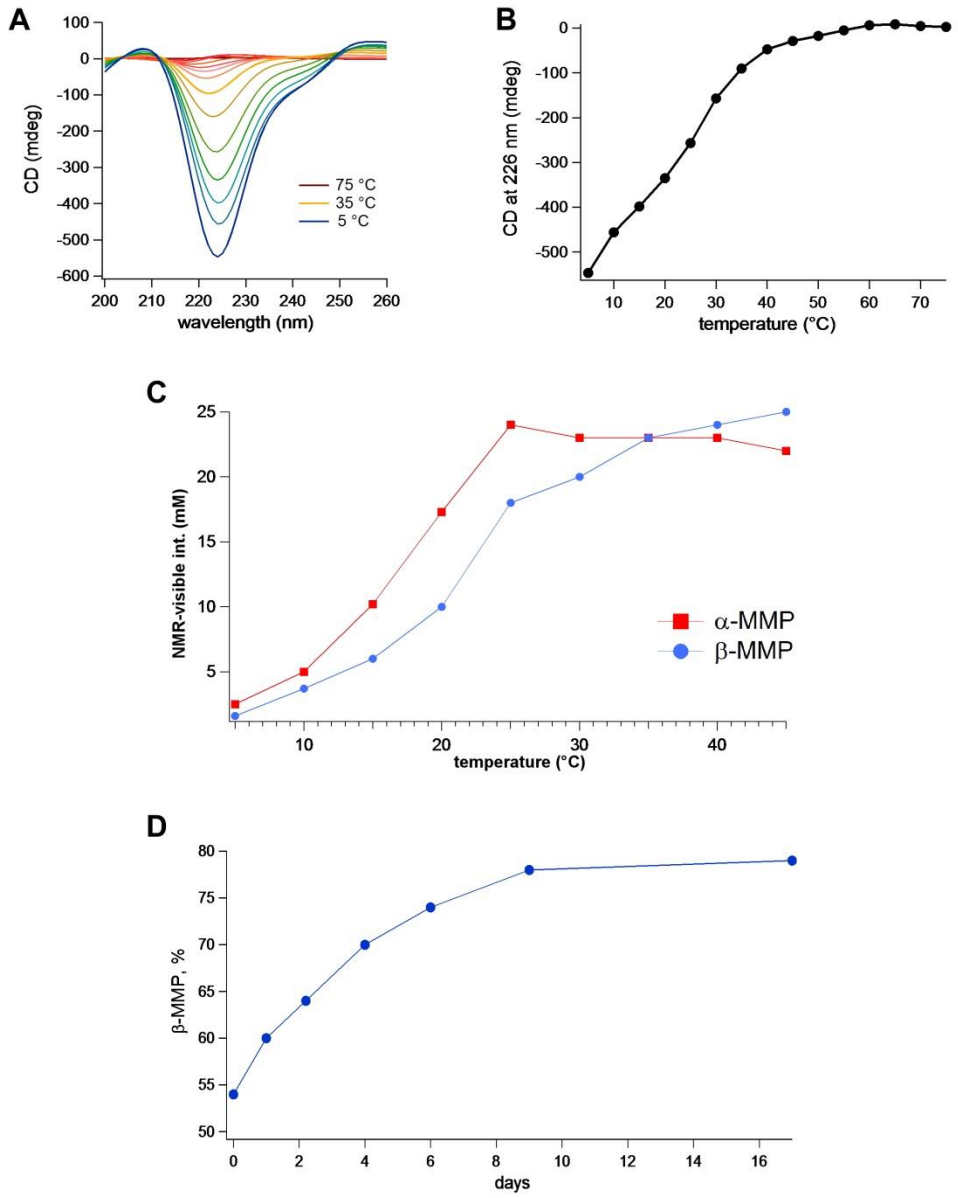


Figure 7.25 CD and ^{1}NMR analysis of MMP assembly with Cy and anomerization in water. (A) Variable temperature CD analysis of a solution containing MMP and Cy at temperatures between 5 to 75 °C. (B) Plot derived from spectra in panel A showing the change in CD intensity at 226 nm as a function of temperature. (C) Plot showing amount of both anomers assembled with Cy at various temperatures determined by NMR analysis. (D) Plot showing the change in anomeric ratio (by percent of *β*-MMP present in solution) as a function of time in a solution containing MMP and Cy incubated at 20 °C. All solutions were 50 mM in both MMP and Cy and 300 mM in NaCl.

7.4 Conclusion

The data presented here demonstrate prebiotically plausible and high yielding syntheses of complementary nucleotides with melamine and barbituric acid as nucleobases under compatible conditions. Moreover, the tendency for melamine and barbituric acid to self-assemble in aqueous solution through the formation of H-bonded base pairs could have facilitated the co-localization of these nucleotides before the advent of biological mechanisms for localization (e.g., compartmentalization). The formation of these assemblies in the unpurified reaction mixtures demonstrates the robustness of the assemblies as they form even in the presence of various side products and staring materials. For both BMP and MMP the β-anomer of the nucleotide is selected in favor of the α-anomer, either as a direct product of its synthesis, as shown for BMP formation, or as a change in the equilibrium distribution due to the presence of supramolecular assemblies, as demonstrated for β-MMP. As β-furanose is the sugar form found in extant nucleic acids, the outcome of the prebiotic reaction, or the ability of the assemblies to discriminate between anomers and promote conversion to the β-form, provides two mechanisms by which nucleotide sugar structure may have been first selected. Furthermore, as MMP and BMP structurally resemble two of the nucleotides found in life today (AMP and UMP, respectively) and have been reported to pair with extant, complementary nucleobases [25, 26], it is tempting to speculate that these heterocycles could have been ancestors of the contemporary genetic alphabet [27].

7.5 References

1. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis. VI. synthesis of purine nucleosides.* J Mol Biol, 1972. **67**: p. 25-33.

2. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis: VII. solid-state synthesis of purine nucleosides.* J Mol Evol, 1972. **1**: p. 249-57.

3. Powner, M.W., B. Gerland, and J.D. Sutherland, *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions.* Nature, 2009. **459**: p. 239-242.

4. Orgel, L.E., *Prebiotic chemistry and the origin of the RNA world.* Crit Rev Biochem Mol Biol, 2004. **39**: p. 99-123.

5. Cafferty, B.J., Hud, N.V., *Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology.* Curr Opin Chem Biol, 2014. **22**: p. 146-157.

6. Joyce, G.F., et al., *The case for an ancestral genetic system involving simple analogs of the nucleotides.* Proc Natl Acad Sci U S A, 1987. **84**(13): p. 4398-402.

7. Hud, N.V., et al., *The origin of RNA and 'My Grandfather's Axe'.* Chem Biol, 2013. **20**: p. 466-474.

8. Kolb, V.M., J.P. Dworkin, and S.L. Miller, *Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides.* J Mol Evol, 1994. **38**: p. 549-557.

9. Bean, H.D., et al., *Formation of a β-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction.* J Am Chem Soc, 2007. **129**: p. 9556-9557

10. Chen, M.C., et al., *Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle.* J Am Chem Soc, 2014. **136**: p. 5640-5646.

11. Menor-Salvan, C., et al., *Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases.* Chem Eur J, 2009. **15**: p. 4411-4418.

12. Whitesides, G.M., et al., *Noncovalent Synthesis: Using Physical-Organic Chemistry To Make Aggregates.* Acc Chem Res, 1995. **28**: p. 37-44.

13. Hysell, M., J.S. Siegel, and Y. Tor, *Synthesis and stability of exocyclic triazine nucleosides.* Org Boimol Chem, 2005. **3**: p. 2946-2952.

14. van Vliet, M.J., J. Visscher, and A.W. Schwartz, *Hydrogen-bonding in the template-directed oligomerization of a pyrimidine nucleotide analog.* J Mol Evol, 1995. **41**: p. 257-261.

15. Kim, H.J. and S.A. Benner, *Prebiotic glycosylation of uracil with electron-donating substituents.* Astrobiology, 2015. **15**: p. 301-306.

16. Krishnamurthy, R., S. Guntha, and A. Eschenmoser, *Regioselective alpha-Phosphorylation of Aldoses in Aqueous Solution.* Angew Chem Int Ed, 2000. **39**: p. 2281-2285.

17. Rios, A.C. and Y. Tor, *On the Origin of the Canonical Nucleobases: An Assessment of Selection Pressures across Chemical and Early Biological Evolution.* Israel Journal of Chemistry, 2013. **53**: p. 469-483.

18. Wulff, G. and G. Clarkson, *On the synthesis of C-glycosyl compounds containing double-bonds without the use of protecting groups.* Carbohydr Res, 1994. **257**: p. 81-95.

19. Hamma, T. and A.R. Ferre´-D'Amare´, *Pseudouridine Synthases.* Chem Biol, 2006. **13**: p. 1125-1135.

20. Zhang, H.C. and G.D. Daves, *Syntheses of 2'-deoxypseudouridine, 2'-deoxyformycin B, and 2',3'-dideoxyformycin B by palladium-mediated glycal-aglycon coupling.* J Org Chem, 1992. **57**: p. 4690-4696.

21. Pankiewicz, K., A. Matsuda, and K.A. Watanabe, *Nucleosides. 121. Improved and general synthesis of 2'-deoxy-C-nucleosides. Synthesis of 5-(2-deoxy-.beta.-D-erythro-pentofuranosyl)uracil, -1-methyluracil, -1,3-dimethyluracil, and -isocytosine.* J Org Chem, 1982. **47**: p. 485-488.

22. Besenius, P., et al., *Controlling the growth and shape of chiral supramolecular polymers in water.* Proc Natl Acad Sci U S A, 2010. **107**: p. 17888-17893.

23. Hirst, A.R., et al., *Low-molecular-weight gelators: elucidating the principles of gelation based on gelator solubility and a cooperative self-assembly model.* J Am Chem Soc, 2008. **130**: p. 9113-9121.

24. Cafferty, B.J., et al., *Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues.* J Am Chem Soc, 2013. **135**: p. 2447-2450.

25. Xia, X., X. Piao, and D. Bong, *Bifacial Peptide Nucleic Acid as an Allosteric Switch for Aptamer and Ribozyme Function.* J Am Chem Soc, 2014. **136**: p. 7265-7268.

26. Bolz, I. and S. Spange, *An enolisable barbiturate with adjustable hydrogen-bonding structure for UV/Vis detection of nucleic acid bases and related compounds.* Chemistry, 2008. **14**: p. 9338-46.

27. Cafferty, B.J. and N.V. Hud, *Was a Pyrimidine-Pyrimidine Base Pair the Ancestor of Watson-Crick Base Pairs? Insights from a Systematic Approach to the Origin of RNA.* Isr J Chem, 2015. DOI: 10.1002/ijch.201400206.

CHAPTER 8

INTERCALATOR MEDIATED ASSEMBLY OF NON-CANONICAL NUCLEIC ACIDS[g]

8.1 Introduction

Challenges with non-enzymatic template directed RNA polymerization using activated racemic nucleotides (D/L ribose) first motivated Joyce and Orgel to propose that RNA was preceded by pre-RNA with more flexible, structurally simpler backbones. 1, 2]. This proposal circumvents problems when incorporating ribose abiotically into RNA or a pre-RNA nucleotides and polymers, problems including: poor yielding prebiotic model reactions to form ribose (which is formed among many other sugars) [3], the inability faithfully copy RNA templates when both chiral forms of a given nucleotide are present (which otherwise results in enantiomeric cross-inhibition, as was first reported by Joyce and Orgel) and connectivity issues (2',5'-linkages would be as common as 3',5'-linkages) [4, 5]. As we have discussed in the preceding chapters, several heterocycles related to the natural nucleobases overcome challenges with nucleoside formation and selection; however it is also important to consider nucleic acid backbones that are prebiotically accessible and capable of rudimentary function. It is widely included among the criteria for a polymer to be considered a pre-RNA candidate that the polymer is able to form stable base paired assemblies with itself and to cross-pair with RNA, or an earlier ancestor of RNA, to enable a continuum of information transfer over the course of nucleic acid evolution [6, 7].

g. Work presented in this chapter was performed in collaboration with Dr. Caterina Musetti.

Attention has primarily focused on pre-RNA candidates that are chemically simpler than RNA, and possess physical properties that may provide an advantage to their abiotic formation or replication [8]. One of the first pre-RNA backbones proposed was a polymer with ribose replaced by a glycerol moiety, which is connected to the nucleobase through a methylene linkage [1]. This alternative nucleic acid, since renamed FNA for flexible nucleic acid (Figure 8.1), was initially attractive as the flexibility of the backbone was theorized to be able to overcome several problems including enantiomeric cross-inhibition. However, enthusiasm for FNA as a pre-RNA candidate declined after it was shown that this nucleic acid does not form stable duplexes [9]. Another glycerol derived nucleic acid was later proposed, glycol nucleic acid (GNA) (Figure 8.1), which showed strong self-pairing properties [10]. However, cross-pairing with the natural nucleic acids was limited to RNA with certain sequences (adenosine and thymidine only) [11]. In a recent publication, Kerry *et. al.* presented a new glycerol-derived nucleic acid, isoGNA, a structural isomer of GNA that differs in the connectivity of its glycol moiety (Figure 8.1). The base-pairing properties of isoGNA were evaluated and assembly was found to be highly sequence dependent with ig$(A)_n$ pairing with ig$(T)_n$ sequences, but no duplex formation was observed when self-complimentary ig$(AT)_n$ was evaluated. Moreover, cross-pairing with the corresponding complementary RNA or DNA sequences was possible for ig$(A)_n$ but not for ig$(T)_n$ sequences or the mixed sequence variants.

Hud and Anet proposed that small aromatic molecules, termed *molecular midwives*, may have facilitated the formation and evolution of informational polymers [12]. Using nucleic acid intercalators, our lab has previously shown that they can aid in non-enzymatic polymerization of short sequences by promoting the stability of duplexes

containing short oligomers [13, 14]. Intercalation mediated assembly has also been shown to promote some tertiary DNA structures [15], and even enable the formation of non-Watson Crick paired duplexes [16]. These results indicate that intercalators can enable programmable assembly, paving the way for their use in nanotechnology [17].

While intercalation of the 2',5'-RNA backbone has been demonstrated [18, 19], the ability of intercalators to promote/enable assembly of non-canonical nucleic acids has not. Here we show that a well-characterized nucleic acid intercalator, proflavine, can bind to isoGNA duplexes with an unprecedented dissociation constant and drive the sequence specific assembly of otherwise non-pairing nucleic acid duplexes. Additionally, proflavine mediated assembly of 2',5'-RNA-DNA duplexes (nucleic acids that assemble poorly) [20] and unlocked nucleic acids (UNA, a backbone related to FNA, which has no homoduplex or cross-pairing ability) [21] were investigated. These findings suggest that the structural landscape of polymers that can facilitate information transfer may be expanded if pairing systems are investigated in the presence of intercalative molecules. Furthermore, our results provide insight into the molecular basis of intercalation beyond previous studies that have only focused on natural nucleic acids by extending the investigation to alternative nucleic acid systems that lack a sugar cycle.

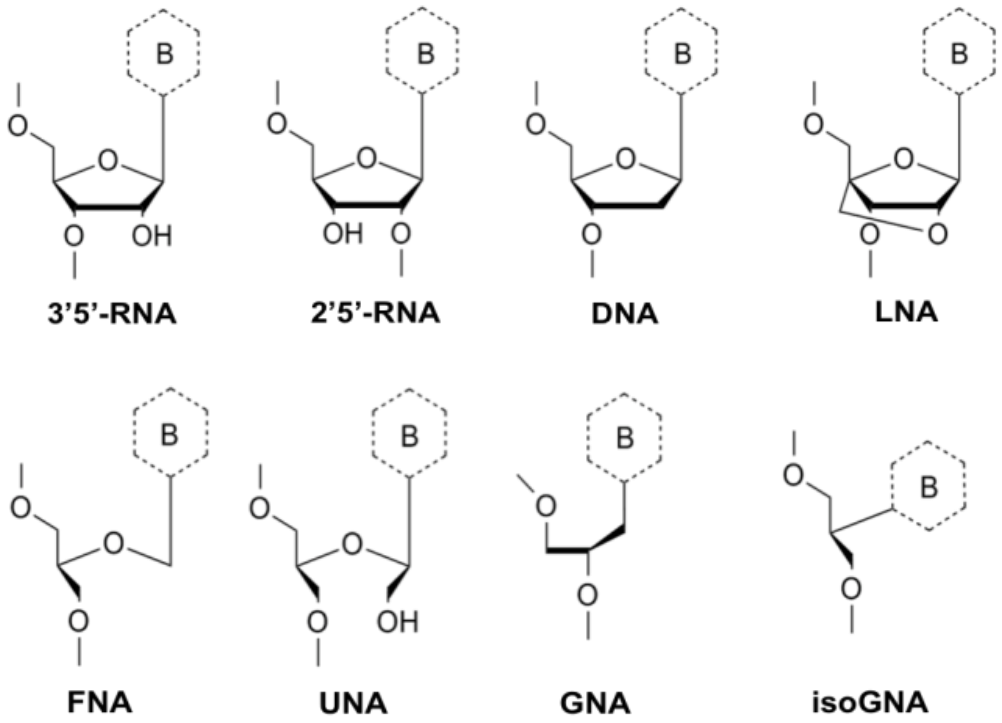


Figure 8.1 Several nucleic acid backbones that have been previously studied for their ability to form duplex assemblies. B indicates a nucleobase attached.

8.2 Experimental Procedures

8.2.1 Materials

Lyophilized synthetic DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), HPLC purified and characterized by LCMS. IsoGNA was synthesized as reported [22, 23]. Due to degradation, all isoGNA oligomers were terminated with a 1'-O- and a 3'-O- DNA nucleotide. 2',5'-RNA oligonucleotides were synthesized in house on a 1 μmol scale with an automated Expedite synthesizer using standard phosphoramidite (Adenosine (n-bz) 3'-tBDSilyl CED phosphoramidite and Uridine 3'-tBDSilyl CED phosphoramidite) from ChemGenes (Wilmington, MA) and standard RNA deprotection. UNA oligonucleotides were purchased from Link Technologies (Bellshill, Scotland) and synthesized in house on a 0.25 μmol scale with an automated Expedite synthesizer using standard phosphoramidite chemistry.

Prior to each experiment, duplex forming sequences were dissolved in buffer (1M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0) to the desired concentration, heated at 90 °C for 5 minutes and cooled slowly to insure the annealing of the two strands.

Table 8.1 Nucleic acid sequences and corresponding abbreviations used in the cross-pairing study. See Table 8.2 for sequences used in the mixed sequence 2',5'-RNA study.

dT-isoGNA(T)$_{16}$-dT	dT-igTTT-TTT-TTT-TTT-TTT-T-dT
dA-isoGNA(A)$_{16}$-dA	dA-igAAA-AAA-AAA-AAA-AAA-A-dA
isoGNA(AT)$_8$	dA-igATA-TAT-ATA-TAT-ATA-T-dT
d(A)$_{18}$	dAAA-AAA-AAA-AAA-AAA-AAA
r(A)$_{18}$	rAAA-AAA-AAA-AAA-AAA-AAA
d(TA)$_8$	dT-dTAT-ATA-TAT-ATA-TAT-A-dA
2',5' RNA(A)$_{18}$	rAAA-AAA-AAA-AAA-AAA-AAA
2',5' RNA(U)$_{18}$	rUUU-UUU-UUU-UUU-UUU-UUU
UNA(A)$_{18}$	rAAA-AAA-AAA-AAA-AAA-AAA
UNA(A)$_{18}$	rUUU-UUU-UUU-UUU-UUU-UUU
mm2C	dAAA-AAC-AAA-AAA-CAA-AAA
mm3C	dAAA-ACA-AAA-CAA-ACA-AAA
mm4C	dAAC-AAA-CAA-ACA-AAC-AAA
mm5C	dAAC-AAC-AAC-AAC-AAC-AAA

Proflavine hemisulfate was purchased from Sigma-Aldrich and used as received. The extinction coefficient used to determine stock solution concentration of proflavine was $\varepsilon_{444} = 38{,}900\ M^{-1}\ cm^{-1}$.

8.2.2 Thermal Denaturation Studies

Thermal denaturation studies were performed on a Cary 300 Bio UV-Vis spectrophotometer in quartz cuvettes of 1 mm and 2 mm path-length. Solutions of nucleic acid (90 μM bp concentration) were combined with proflavine at 45 μM concentration. UV melting profiles were produced by monitoring absorbance at 260 nm while increasing the temperature at a rate of 0.4 °C min^{-1} from 0 °C to 65 °C or from 0 °C to 90 °C. Recordings were taken during both the melting and annealing process. The melting temperature values (T_m) were determined from the first derivative of the melting profiles.

8.2.3 Dilution Studies

Dilution studies were conducted at 4 °C and 20 °C on an Agilent UV-Vis spectrophotometer equipped with an Agilent 89090A Peltier temperature controller. A buffer solution (1M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0) was added incrementally to a sample containing a constant ratio of 2:1 (nucleic acid:proflavine). Initial concentration was 90 μM in oligonucleotide base-pair and 45 μM of proflavine. A rectangular quartz cuvette with a 10 mm pathlength and cylindrical cuvettes with a 5 or 10 cm pathlength (Starna Cells) were used depending on concentration to maintain an absorbance below 1.2 AU. Spectra collected were then normalized to extinction coefficient in units $M^{-1}cm^{-1}$.

8.2.4 Circular Dichroism (CD) Analysis.

CD spectra were acquired at 4 °C and at 20 °C on a JASCO J-810 CD spectropolarimeter equipped with a NESLAB temperature controller. CD spectra were recorded at 4 °C from 220 to 600 nm using a 1 mm and 10 mm path-length cell. All the reported spectra were baseline corrected with a separately acquired buffer spectrum. Each spectrum represents the average of two scans recorded with 1 nm step resolution and 4 sec response time.

8.3 Results and Discussion

8.3.1 Binding of Proflavine to an isoGNA Duplex is Consistent with Intercalation

The base-pairing properties of two complementary isoGNA homogeneous sequences, dT-isoGNA$(T)_{16}$-dT and dA-isoGNA$(A)_{16}$-dA were first investigated under conditions reported by Karri et al. (1M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0) [24]. Consistent with earlier results, the complementary isoGNA strands form a duplex with a T_m of 48 °C (Figure 8.2A). A large hysteresis of 39 °C was observed under the conditions of the melt experiment (approx. 2 °C per minute) indicating a loss in single strand preorganization and potentially alternative intra-strand interactions. This hysteresis may result from the ig(T) strand which would be predicted to have poorer intra-strand base stacking than ig(A) when compared to analogous DNA melts. The high salt concentration was necessary, likely to screen the phosphate charges between the strands which would be predicted to be more repulsive due to the nucleobase being directly attached to the isoGNA backbone resulting in a closer inter-strand distance (for comparison, the nucleobase is 3 bonds away in RNA and DNA).

Upon addition of proflavine at a ratio of one intercalator per two base pairs (nearest neighbor-exclusion loading) hysteresis decreased by 11 °C and the stability of the iso(T)-iso(A) duplex increased by 3 °C (Figure 8.2B). These results indicate favorable interactions between proflavine and the isoGNA duplex. Moreover, the UV-vis spectra of unassembled proflavine relative to proflavine complexed with the isoGNA duplex shows a red shift in the absorption band from 444 nm to 447 nm (Figure 8.2C), this spectral feature, also seen upon intercalation of proflavine or ethidium into DNA duplexes, suggests that binding occurs through an intercalative mode.

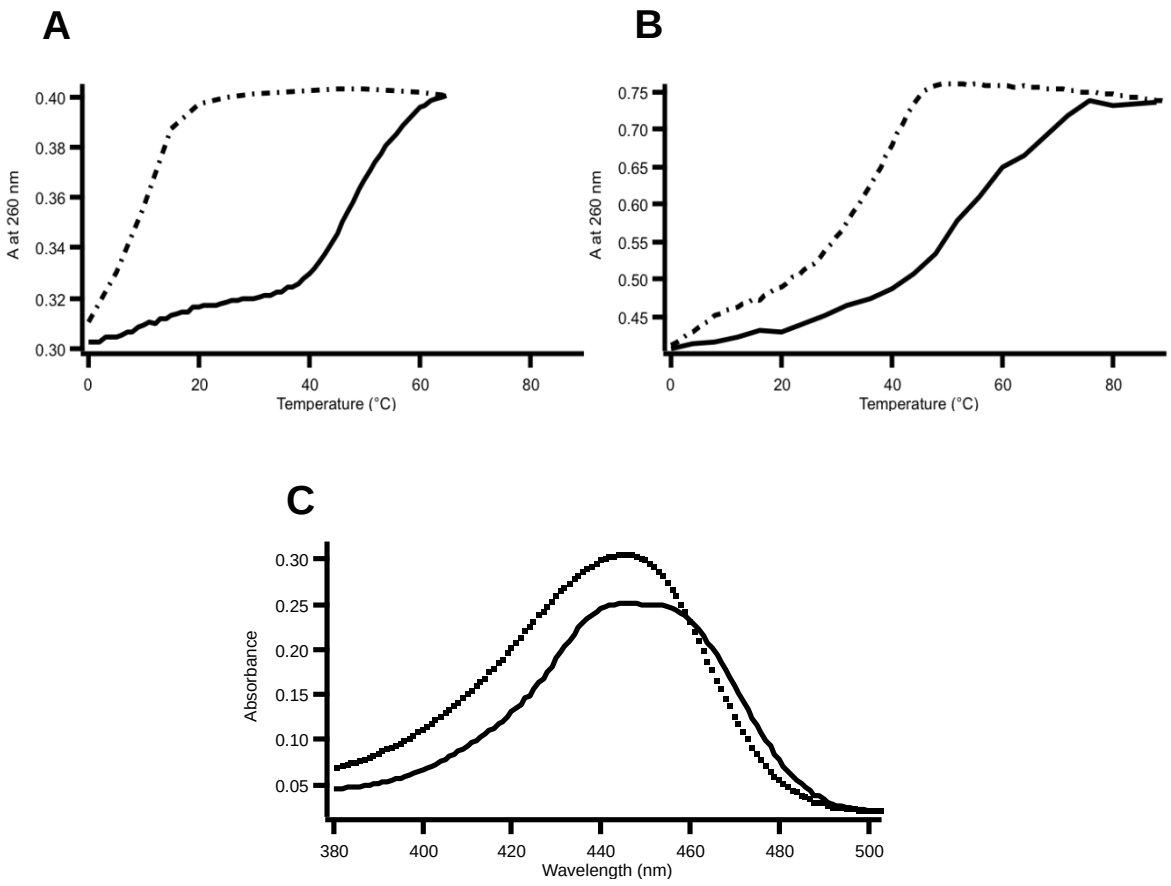


Figure 8.2 Melting profiles of dT-isoGNA$(T)_{16}$-dT/dA-isoGNA$(A)_{16}$-dA in the (A) absence, and in the (B) presence of proflavine. (C) Proflavine UV-vis absorbance region in the absence (dotted line) or presence (solid line) of the isoGNA duplex dT-isoGNA$(T)_{16}$-dT/dA-isoGNA$(A)_{16}$-dA. Sample solutions contained 1M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0.

Nearest neighbor-exclusion loading concentrations (that is one intercalator per two base pairs) were initially used as this is the upper limit for the number of intercalators that can bind DNA or RNA duplexes, a property of nucleic acid intercalation known as the nearest neighbor exclusion principle (NNEP) [25]. Circular dichroism (CD) was next employed to obtain a better understanding of proflavine's binding mode. While the monomeric unit of the isoGNA backbone is achiral, both dT-isoGNA$(T)_{16}$-dT and dA-isoGNA$(A)_{16}$-dA contain one chiral D-deoxynucleotide at each strand terminus, providing both molecular and supramolecular (duplex) chirality. An induced band in the absorption region of achiral proflavine was observed when tested at nearest neighbor loading concentrations, indicating that proflavine is binding to the duplex. To further evaluate the stoichiometry of binding, a concentrated proflavine solution was titrated into a solution containing a set concentration of the isoGNA duplex. As shown in Figure 8.3, a small CD intensity is observed for both isoGNA duplex that are not in the presence of proflavine (unbound), as well as samples containing a low concentration of proflavine. However, increasing the concentration of ligand at and beyond a stoichiometery of one proflavine per two base pairs generated a large CD signal in both the nucleic acid and proflavine absorption regions. A similar results was obtained with Job's method of continues variation [26], which shows an inflection point at nearest neighbor-exclusion loading (Figure 8.3). These results are consistent with intercalative binding and adherence to the NNEP.

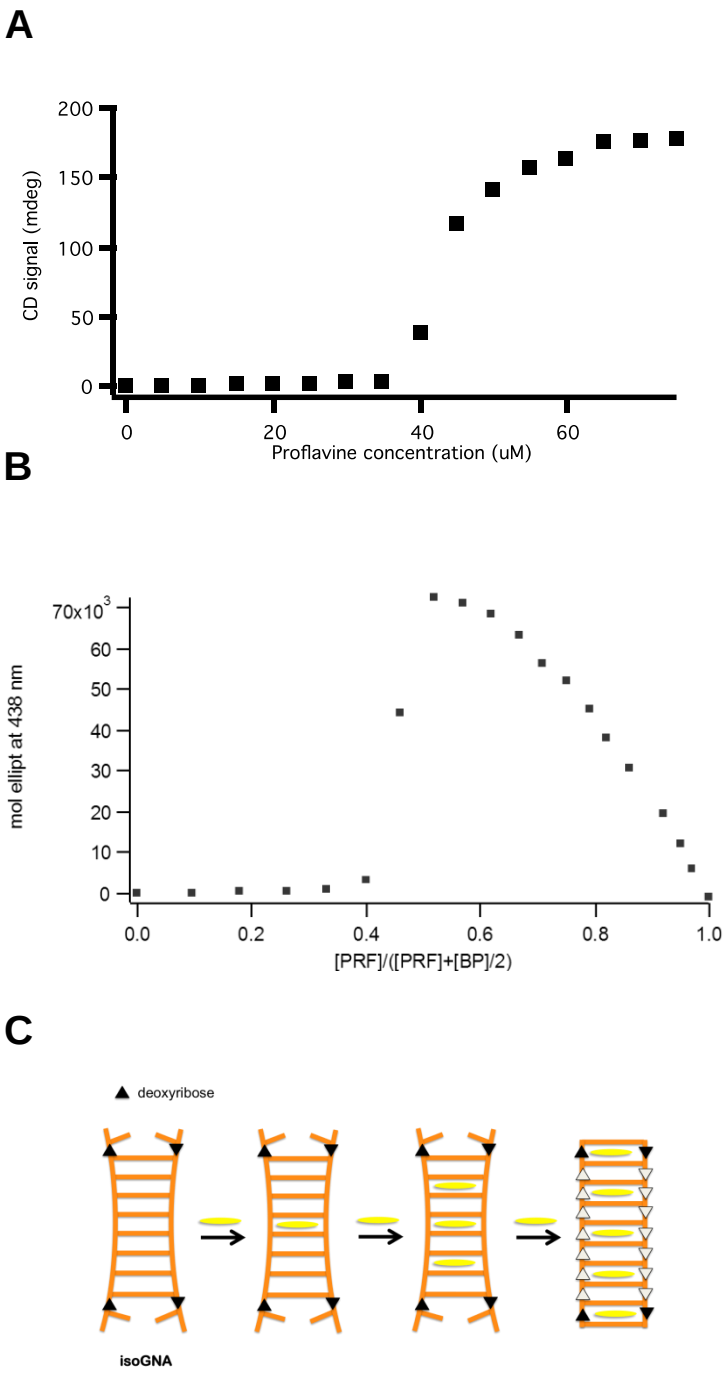


Figure 8.3 Stoichiometry of proflavine binding to isoGNA duplex determined by CD analysis. (A) Titration of the dT-isoGNA$(T)_{16}$-dT/dA-isoGNA$(A)_{16}$-dA duplex (90 in base pair) with proflavine. (B) Job plot analysis of the binding of proflavine to the dT-isoGNA$(T)_{16}$-dT/dA-isoGNA$(A)_{16}$-dA duplex. (C) Model for intercalator loading of the isoGNA sequence which explains the sudden increase in CD signal upon full loading which occurs at 45 um in A and 0.5 in B. Briefly, the low intensity of the duplex CD signal due to poor chiral propagation from the only chiral groups on the oligonucleotides, i.e., DNA nucleosides (black triangles) at the terminals. Proflavine loads first at the most favorable positions and finally at the terminal base pairs, locking the chiral nucleotides in position and providing the corresponding CD signal.

CD experiments suggest that proflavine loading occurs progressively, starting from the center of the duplex and moving outward, as well as simultaneously among the entire duplex population (Figure 8.3C). The small CD signal at lower proflavine:base pair ratios can be understood as proflavine would initially bind to the most energetically favorable sites at the center of the duplex, away from the chiral termini that would be less ordered due to end fraying. Upon increasing the ratio of proflavine, loading would occur from the center of the nucleic acid to the external base pairs. At nearest neighbor-exclusion loading concentration, the external base pairs are stabilized by the ligand which propagates the chiral signal throughout the fully loaded proflavine bound duplex.

We note that adherence to NNEP for proflavine binding to an isoGNA duplex was initially only speculative as NNEP's origin is still debated. Previously, it has been suggested that the NNEP results from steric constraints caused by conformational changes of the sugar upon addition of an intercalator or due to the inability of the backbone to stretch within a duplex to enable an intercalator to bind above the concentration determined by NNEP [25]. However, molecular dynamics (MD) results suggest that the DNA backbone is flexible enough to accommodate full loading [27, 28]. Experimental investigations into this proposal are largely lacking as all previous reports on intercalation involve nucleic acids that contain the ribose sugar cycle [19, 29]. Because isoGNA lacks a sugar cycle (and thus would not suffer from the same steric considerations, e.g., sugar pucker) and still conforms to NNEP it is unlikely that pure steric exclusion is the determining factor. Therefore, these findings provide further insight into the mechanism governing the nearest neighbor exclusion principle.

8.3.2 Submicromolar Binding of Proflavine to an isoGNA Duplex

Dilution studies were next performed to investigate proflavine's binding affinity to an isoGNA duplex. As stated above, combining ig(A), ig(T) and proflavine results in a red shift of the absorption maxima of proflavine from 444 nm to 447 nm, characteristic of the unbound and bound states of the small molecule, respectively (Figure 8.2). This spectral feature was used to determine the apparent K_d of the proflavine-isoGNA assembly by monitoring the progressive shift in proflavine's absorbance maxima, toward that of the spectrum corresponding to free proflavine, upon titration with buffer. An apparent K_d of 0.4 μM was determined for the proflavine bound dT-isoGNA$(T)_{16}$-dT-dA-isoGNA$(A)_{16}$-dA homoduplex (Figure 8.4). The affinity of proflavine for this isoGNA duplex is extremely strong, one order of magnitude higher than with natural DNA (K_d= 12 μM) under the same ionic strength and temperature conditions (4 °C).

The enhanced binding of proflavine to the isoGNA duplex compared with the DNA analog may result from: 1) the more flexible isoGNA backbone adopting a more favorable conformation at the intercalation site; 2) would not be hindered by steric factors related to the presence of the sugar cycle of ribose in the DNA/RNA systems (e.g., sugar pucker); 3) better charge pairing interactions between the positively charged proflavine and more closely spaced phosphate groups of isoGNA. Under these conditions, this low apparent K_d is intriguing as it is known that high salt concentrations (e.g., 1M NaCl) will shield electrostatic interactions between nucleic acids and charged intercalators thus thereby reducing binding affinity [19].

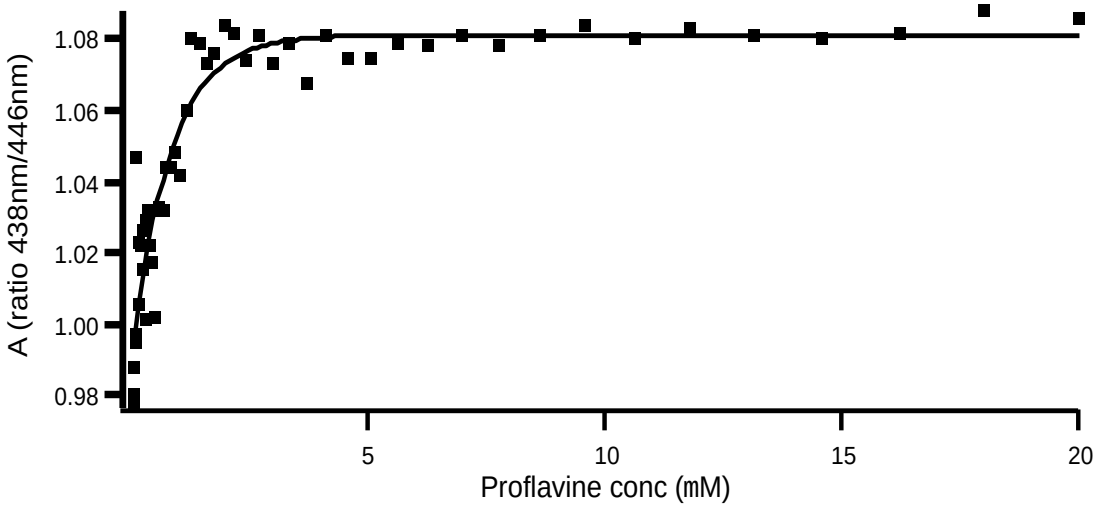


Figure 8.4 Dilution study used to determine an apparent binding constant for proflavine to the dT-isoGNA$(T)_{16}$/dA-isoGNA$(A)_{16}$-dA_8 complex. Solution contained isoGNA and proflavine at a ratio of one proflavine molecule per two base pairs of the isoGNA duplex (nearest neighbor-exclusion loading). At 4 °C the apparent K_d was determined to be 0.4 mM. Buffer contained 1M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0.

8.3.3 Proflavine Enables Formation of an isoGNA Duplex

We next investigated the ability of proflavine to bind to a self-complementary isoGNA sequence consisting of alternating A and T nucleotides (dT-isoGNA$(AT)_8$-dA). This heterogeneous sequence was previously reported to be unable to form base paired assemblies [24]. This prompted us to investigate if proflavine can promote the assembly of an otherwise non-pairing isoGNA sequence. While no transition was observed when proflavine was omitted, the melting profiles of the mixed isoGNA sequence in the presence of proflavine show a cooperative transition, with a T_m of 37 °C (Figure 8.5). This finding demonstrates that otherwise non-assembling alternative nucleic acids can form a stable duplex via noncovalent interactions resulting from small molecule binding. Noteworthy is the absence of hysteresis during the duplex re-annealing process (i.e., cooling trace) which was observed for the ig(A)/ig(T) system in both the presence and absence of an intercalator. This difference may result from an increase in internal structure of the mixed sequence oligonucleotide, as sequences containing purines will possess more internal order due to enhanced stacking interactions when compared to sequences rich in pyrimidines, such as the ig(T) sequence [30].

The apparent binding constant of proflavine for the dT-ig$(AT)_8$-dA duplex was determined to be 1 μM (Figure 8.5C). The observed lower affinity of proflavine for the mixed sequence duplex compared to the ig(A)-ig(T) system can be rationalized due to the additional binding energy needed to stabilize duplex formation. Even so, the K_d for the proflavine bound dT-ig$(AT)_8$-dA duplex is still lower than for proflavine bound to a preassembled DNA duplex under the conditions tested for this study.

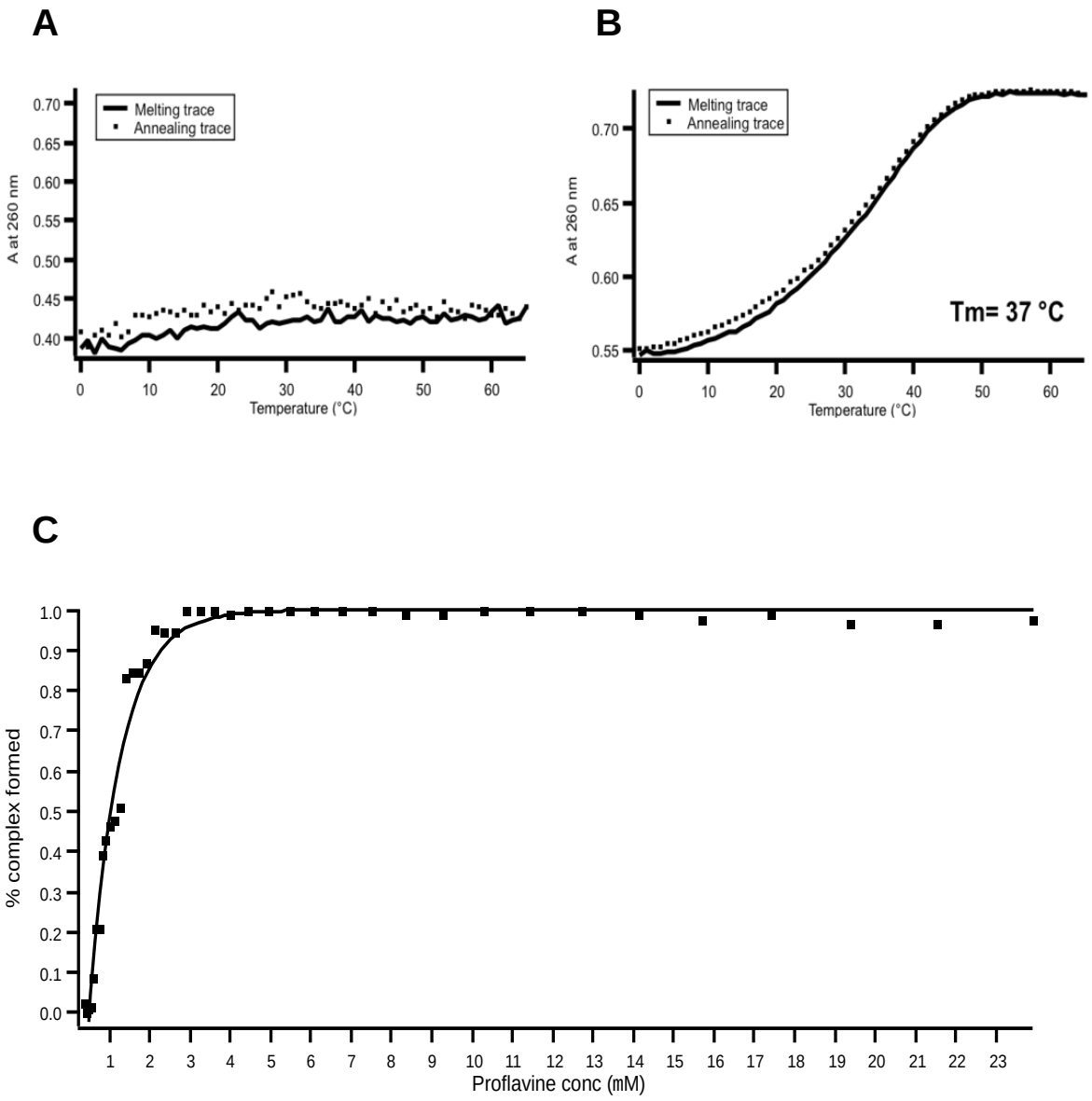


Figure 8.5 UV-vis analysis used to determine melting temperatures and binding constants for proflavine/dT-isoGNA(AT)$_8$-dA assemblies. Melting (solid line) and annealing (dotted line) profiles of the dT-isoGNA(AT)$_8$-dA system in the (A) absence, and in the (B) presence of proflavine. (C) Apparent binding constant determination for the proflavine isoGNA(AT)$_8$ complex; at 4 °C the apparent K_d was determined to be 1 mM. All solution contained isoGNA and proflavine at a ratio of one proflavine molecule per two base pairs of the isoGNA duplex (nearest neighbot loading). Buffer conditions for all experiments was 1 M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0.

8.3.4 Proflavine Enables Formation of isoGNA Hetero-Duplexes

The isoGNA sequence dA-isoGNA$(A)_{16}$-dA was reported to cross-pair with complementary DNA and RNA, while dT-isoGNA$(T)_{16}$-dT was unable to pair with either [24]. Given the positive results obtained for assembling the mixed sequence isoGNA above, poflavine's ability to enable the heteroduplex formation of dT-isoGNA$(T)_{16}$-dT with complementary $d(A)_{18}$ and $r(A)_{18}$ strands was next evaluated. The addition of proflavine to solutions containing the complementary but non-pairing DNA or RNA sequences was found to enable the formation of stable heteroduplexes (Figure 8.6). The melting temperature for the isoGNA-DNA heteroduplex was 28 °C, while the isoGN-RNA heteroduplex showed an impressive T_m of 44 °C (Figure 8.6). The large difference between the two heteroduplexes indicates that subtle backbone changes can play a large role in intercalator-mediated assembly.

The sequence specificity of intercalator driven heteroduplex formation was next studied by investigating assemblies of proflavine, dT-isoGNA$(T)_{16}$-dT and partially complementary DNA strands containing from two to five A to C substitutions. A linear decrease in T_m occurred with increasing C:T mismatches for both the isoGNA-DNA heteroduplex and a control $d(T)_{18}$-$d(A)_{18}$ DNA homoduplex, demonstrating that the proflavine mediated assemblies are sequence dependent (Figure 8.6). For the DNA control, the presence of four mismatches was found to inhibit duplex formation even with proflavine. Mismatches were better tolerated for these isoGNA-DNA heteroduplex. In this case, proflavine enabled duplexation in the presence of up to five mismatches. Proflavine's higher affinity for duplexes formed by the conformationally more adaptive isoGNA oligonucleotide may be responsible for the loss in mismatch sensitivity.

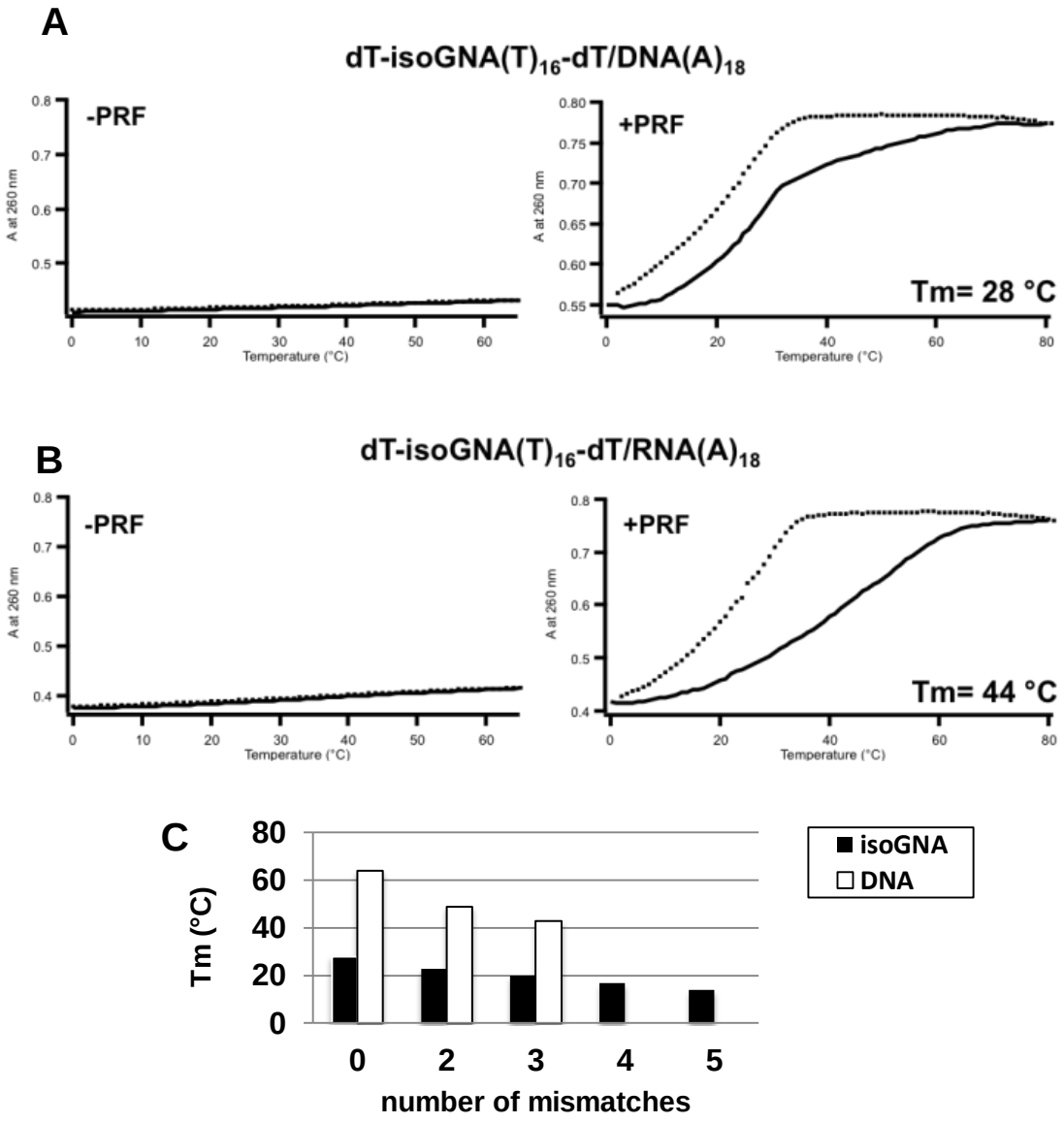


Figure 8.6 Thermal denaturation studies of isoGNA heteroduplexes. Melting (solid line) and annealing (dotted line) profiles of the dT-isoGNA$(T)_{16}$-dT with (A) DNA A_{18}, and (B) RNA A_{18} in the absence (left panels) and in the presence (right panels) of 0.5 equivalents of proflavine per base pair of the isoGNA containing heteroduplex. (C) Melting temperatures obtained with dT-isoGNA$(T)_{16}$-dT (black bars) or the DNA sequence $d(T)_{18}$ (white bars) when combined with proflavine and either a complementary DNA $d(A)_{18}$ or a semi-complementary DNA strand containing two, three, four and five A to C substitutions. Solutions contained 1 M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0.

8.3.5 Intercalation of 2′,5′-RNA Heteroduplexes

In order to evaluate the generality of intercalator assisted non-canonical nucleic acid assembly, we also investigated 2',5'-RNA based systems. The 2',5'-RNA backbone linkage, arguably the closest isomer of canonical 3',5'-linked RNA, has been observed to form during non-enzymatic RNA polymerization [31], and has recently been shown to be tolerated by functional RNAs when incorporated along with 3',5'-linkages [32]. 2',5'-RNA possesses unique, and potentially theraputic, pairing properties since this backbone can hybridize with canonical 3',5'-RNA, however it has been reported not to be compatible with DNA duplex assembly, motiviating investigations into potential antisense therapies [20]. The base-pairing properties of two 2',5'-RNA sequences, $(U)_{18}$ and $(A)_{18}$, and their complementary DNA and RNA strands were examined by UV-vis spectroscopy. As expected, both sequences of 2',5'-RNA pair with complementary 3',5'-RNA, while no interaction was observed for the corresponding 2',5'-RNA/DNA systems (Figure 8.7). The addition of proflavine to the 2',5'-RNA/3',5'-RNA solutions resulted in a small increase (i.e., 2 °C) in T_m and a red shift in the UV absorption region of proflavine. The addition of proflavine into 2',5'-RNA/DNA samples resulted in the appearance of a cooperative and reversible transition at 19 °C for 2′,5′-r$(U)_{18}$/d$(A)_{18}$ and 39 °C for 2',5'-r$(A)_{18}$/d$(T)_{18}$ (Figure 8.7A and B). These T_m values indicate a proflavine dependent stabilization of a 2',5'-RNA/DNA heteroduplex that is *at least* 20 °C, since no transition is observed during thermal denaturation studies. Even more impressive is the T_m of the 2',5'-RNA/isoGNA heteroduplex, which also requires proflavine to assemble, possessing a T_m of 52 °C (Figure 8.7C). This is, to our knowledge, the largest change in thermal stability provided by a small molecule to a nucleic acid duplex.

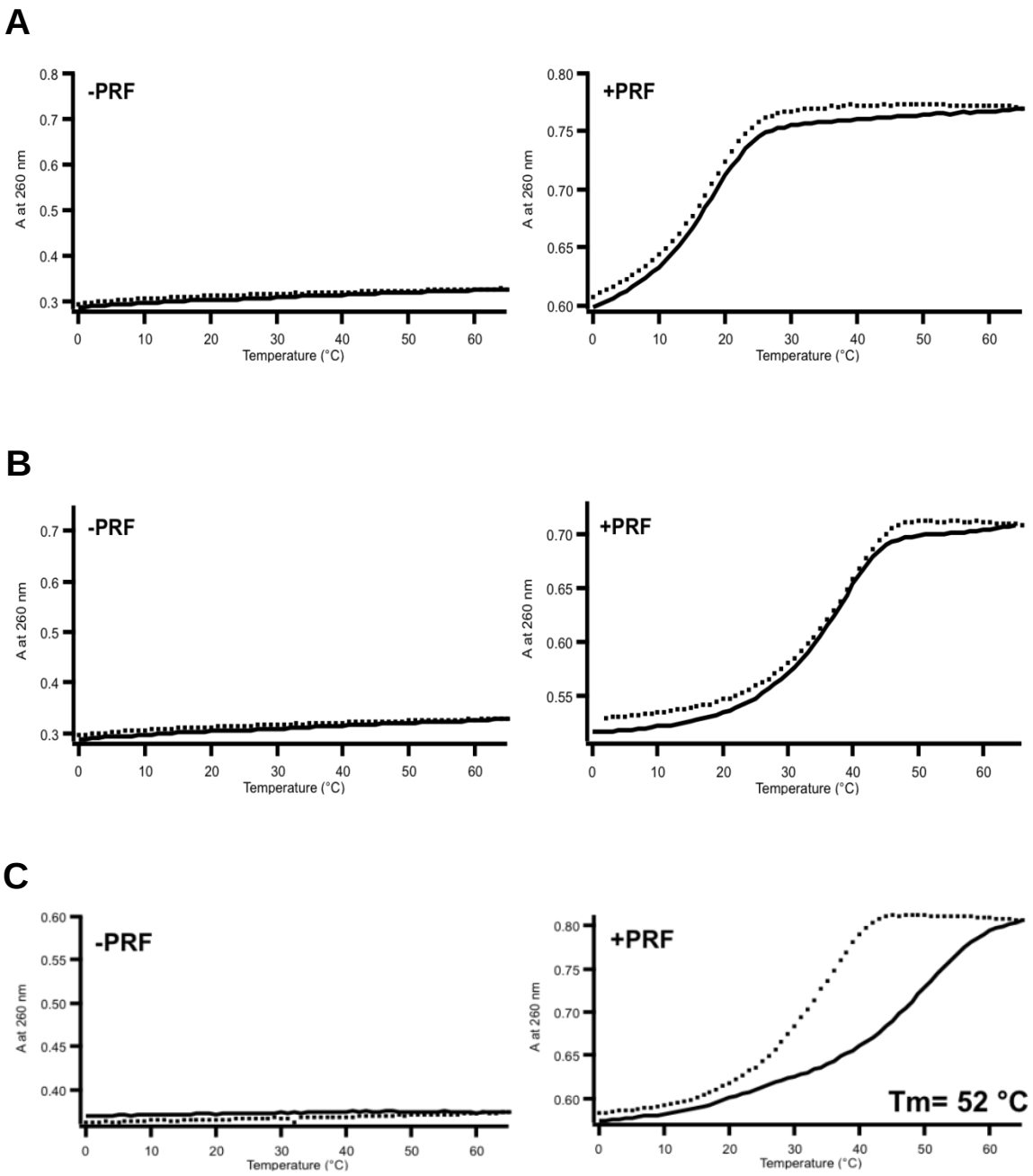


Figure 8.7 Thermal denaturation studies for solutions containing 2',5'-RNA and complementary DNA or isoGNA, in the presence and absence of proflavine. Solutions contained, (A) 2',5'-RNA$(U)_{18}$ and DNA $(A)_{18}$; (B) 2',5'-RNA$(A)_{18}$ and DNA $(T)_{18}$; (C) 2',5'-RNA$(U)_{18}$ and dA-isoGNA$(A)_{16}$-dA_8. Melting traces are shown in solid lines and annealing traces in dotted line. All solutions contained, 1 M NaCl, 10 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0.

To understand how sequence content affects assembly of 2',5'-RNA heteroduplexes, two series of 12-mers were synthesized containing mixed sequences (not self-complementary) with DNA, 3',5'-RNA and 2',5'-RNA backbones. One series contained 100 % GC content (GC series) and the other contained half GC content (ATGC series) (see table 8.2 for sequences). Additionally, sequence mismatches were also evaluated for both series by incorporating semi-complementary sequences within the DNA backbone containing two G to A substitutions. Table 8.3 and Table 8.4 display the average melting temperatures (two heating and cooling traces) obtained from 36 variable temperature UV-vis experiments, analyzing both homoduplexes and heteroduplexes in the presence of absence of proflavine. For these experiments the ionic strength was decreased to 0.1 M NaCl in order to enhance proflavine-duplex assembly.

In contrast to previous reports by Damha and coworkers, we found that a short 2',5'-RNA-DNA duplex will assemble without an intercalator, however, only in the GC series. This observation is not too unexpected, given the full GC sequence content, however, these strands are considerably shorter (12-mers compared to 18-mers) than the systems previously reported. As expected, the 2',5'-RNA/DNA duplex of the ATGC series only assembled in the presence of proflavine. An interesting observation from this comparative study was that the magnitude of the effect of an intercalator to assemble all duplexes was found to diminish with increasing thermal stability of the unintercalated duplex, note that melts above 40°C tend to have a difference in T_m of only a few degrees when in the presence of the intercalator, however, for the 2',5'-RNA/DNA duplex of the ATCG series the increase was greater than 20 °C as no pairing was evident for this system in the absence of intercalator.

Table 8.2 Sequences used in the mixed sequence 2',5'-RNA cross-pairing study.

	GC Series	ATGC Series	Nucleic Acid Type
Seq. 1	5'-CCC **G**CC **G**CG CCG-3'	5'-GTA **G**AC **G**AA TGG-3'	DNA, 3',5'-RNA
mm1 mismatch	5'-CCC **A**CC **A**CG CCG-3'	5'-GTA **A**AC **A**AA TGG-3'	DNA
Seq. 2	5'-CGG CGC GGC GGG-3'	5'-CCA TTC GTC TAC-3'	DNA, 3',5'-RNA, 2',5'-RNA

Table 8.3 Melting temperatures of oligonucleotide duplexes in the GC series. See Table 8.2 for sequences. NM indicates that no melt transition was observed, (T_m<10°C) indicates that a transition was present but no clear baseline was evident to determine the start of the melt. The temperature range was from 5 °C to 95 °C and two cooling and two heating experiments were performed for each system. Each sample contained 100 mM NaCl, 10 mM BPE, 3.33 μM each strand and 20 μM proflavine (+PRF), if applicable.

GC Series	3',5'-DNA 1 (°C)		3',5'-RNA 2 (°C)		3',5'-DNAmm1 (°C)	
	-PRF	+PRF	-PRF	+PRF	-PRF	+PRF
3',5'-DNA 2	66.3	67.5	62.3	66.2	35.8	40.5
3',5'-RNA 2	65.8	65.3	78.3	79.5	41.3	43.8
2',5'-RNA 2	32.8	47.3	65.5	70.8	NM	<10

Table 8.4 Melting temperatures of oligonucleotide duplexes in the ATGC series. See Table 8.2 for sequences. NM indicates that no melt transition was observed, (T_m<10°C) indicates that a transition was present but no clear baseline was evident to determine the start of the melt. The temperature range was from 5 °C to 95 °C and two cooling and two heating experiments were performed for each system. Each sample contained 100 mM NaCl, 10 mM BPE, 3.33 μM each strand and 20 μM proflavine (+PRF), if applicable.

ATGC Series	3',5'-DNA 1 (°C)		3',5'-RNA 2 (°C)		3',5'-DNAmm1 (°C)	
	-PRF	+PRF	-PRF	+PRF	-PRF	+PRF
3',5'-DNA 2	40.8	46.5	40.3	44	TS	22
3',5'-RNA 2	34.5	40.8	51.3	51	NM	<10
2',5'-RNA 2	NM	20	27.8	32.3	NM	<10

8.3.6 Proflavine does not Support Formation of an Unlocked Nucleic Acids Duplex

Both isoGNA and 2′,5′-RNA based duplexes share similar base pairing properties; they are both capable of pairing modes that allow duplex formation but their ability to pair is both sequence and backbone dependent. We next investigated UNA which, in contrast, has not been demonstrated to form homo or heteroduplexes. Thus, this system provided a formidable challenge for intercalator mediated assembly. Additionally, acyclic UNA is very similar to the historically significant FNA backbone described by Joyce and Orgel [1], while also being structurally more similar to RNA when compared to FNA (Figure 8.1).

Two complementary sequences of UNA were synthesized (A18 and U18) and evaluated for their pairing ability together or as heteroduplexes with complementary DNA and RNA. Both UV-vis and CD thermal denaturation studies confirmed that homoduplex formation between the UNA oligonucleotides or heteroduplex formation between UNA and complementary RNA, or DNA does not occur, even in the presence of 1 M NaCl (data not shown). Furthermore, proflavine was not able to rescue duplex formation with any system containing the UNA backbone. The inability of oligo-UNA to form homo- or heteroduplexes may be due to this highly flexible nucleic acid forming intramolecular structures that are more stable than the bimolecular duplex or as a result of preorganization that interferes with the formation of base-paired assemblies (i.e., gauche effect, personal communication with Prof. R. Krishnamurthy). However, additional sequences, buffer conditions and potentially alternative intercalators will first need to be evaluated before a thorough understanding of UNA duplex behavior can be obtained.

8.4 Conclusion

Watson-Crick base pairing is essential for information transfer and processing. Thus, homoduplex formation of an informational polymer and its ability to form heteroduplexes with natural nucleic acids can be considered an essential property for pre-RNA candidacy. The results presented in this chapter highlight the pronounced role of nucleic acid intercalators in overcoming pairing limitations of alternative nucleic acid systems. Furthermore, the inability of proflavine to enable UNA duplex formation demonstrates that the chemical space of nucleic acid backbones that are compatible with asssembly, both intercalator mediated and not, is still constrained. Additionally, the results presented here suggest that modest changes in the nucleic acid backbone architecture could result in drastic binding affinities for intercalators, a finding that may aid in the development of new nucleic acid based technologies.

8.5 References

1. Joyce, G.F., et al., *The case for an ancestral genetic system involving simple analogs of the nucleotides.* Proc Natl Acad Sci U S A, 1987. **84**: p. 4398-402.
2. Joyce, G., et al., *Chiral selection in poly(C)-directed synthesis of oligo(G).* Nature, 1984. **310**: p. 602-604.
3. Decker, P., P. Schweer, and R. Pohlmann, *Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectroscopy of n-butoxime trifluoroacetates on OV-225.* J. Chromatogr., 1982. **244**: p. 281-291.
4. Orgel, L.E., *Prebiotic chemistry and the origin of the RNA world.* Crit Rev Biochem Mol Biol, 2004. **39**: p. 99-123.
5. Lohrman, R. and L.E. Orgel, *Preferential formation of (2'–5')-linked internucleotide bonds in non-enzymatic reactions.* Tetrahedron, 1978. **34**: p. 853-855.

6. Yang, Y.-W., et al., *Experimental evidence that GNA and TNA were not sequential polymers in the prebiotic evolution of RNA.* J Mol Evol, 2007. **65**: p. 289-295.

7. Eschenmoser, A., *Chemical etiology of nucleic acid structure.* Science, 1999. **284**: p. 2118-2124.

8. Engelhart, A.E. and N.V. Hud, *Primitive genetic polymers.* Cold Spring Harbor Perspectives in Biology, 2010. **2**: p. 10.1101/cshperspect.a002196.

9. Schneider, K.C. and S.A. Benner, *Oligonucleotides containing flexible nucleoside analogs.* J Am Chem Soc, 1990. **112**: p. 453-455.

10. Zhang, L.L. and E. Meggers, *An extremely stable and orthogonal DNA base pair with a simplified three-carbon backbone.* J Am Chem Soc, 2005. **127**: p. 74-75.

11. Zhang, L.L., A. Peritz, and E. Meggers, *A simple glycol nucleic acid.* J Am Chem Soc, 2005. **127**: p. 4174-4175.

12. Hud, N.V. and F.A.L. Anet, *Intercalation-mediated synthesis and replication: a new approach to the origin of life.* J Theor Biol, 2000. **205**: p. 543-562.

13. Jain, S.S., et al., *Enzymatic behavior by intercalating molecules in a template-directed ligation reaction.* Angew Chem Int Ed, 2004. **43**: p. 2004-2008.

14. Horowitz, E.D., et al., *Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world.* Proc Natl Acad Sci U S A, 2010. **107**: p. 5288-5293.

15. Greschner, A.A., K.E. Bujold, and H.F. Sleiman, *Intercalators as molecular chaperones in DNA self-assembly.* J Am Chem Soc, 2013. **135**: p. 11283-8.

16. Persil, Ö., et al., *Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding.* J Am Chem Soc, 2004. **126**: p. 8644-8645.

17. Persil, Ö. and N.V. Hud, *Harnessing DNA intercalation.* Trends in Biotechnology, 2007. **25**(10): p. 433-436.

18. Horowitz, E.D., et al., *Solution structure and thermodynamics of 2',5' RNA intercalation.* J Am Chem Soc, 2009. **131**: p. 5831-5838.

19. Horowitz, E.D. and N.V. Hud, *Ethidium and proflavine binding to a 2',5'-linked RNA duplex.* v, 2006. **128**: p. 15380-15381.

20. Giannaris, P.A. and M.J. Damha, *Oligoribonucleotides containing 2',5'-phosphodiester linkages exhibit binding selectivity for 3',5'-RNA over 3',5'-ssDNA.* Nucleic Acids Res, 1993. **21**: p. 4742-4749.

21. Langkjaer, N., A. Pasternak, and J. Wengel, *UNA (unlocked nucleic acid): a flexible RNA mimic that allows engineering of nucleic acid duplex stability.* Bioorg Med Chem, 2009. **17**: p. 5420-5425.

22. Meher, G., et al., *Microwave-assisted preparation of nucleoside-phosphoramidites.* Chem Commun, 2014. **50**: p. 7463-7465.

23. Efthymiou, T. and R. Krishnamurthy, *Microwave-assisted phosphitylation of DNA and RNA nucleosides and their analogs.* Curr Protoc Nucleic Acid Chem, 2015. **60**: p. 2 19 1-2 19 20.

24. Karri, P., et al., *Base-pairing properties of a structural isomer of glycerol nucleic acid.* Angew Chem Inter Ed, 2013. **52**: p. 5840-5844.

25. Voet, D., *Intercalation complexes of DNA.* Nature, 1977. **269**: p. 285-286.

26. Job, P., *Studies on the formation of complex minerals in solution and on their stability.* Annales De Chimie France, 1928. **9**: p. 113-203.

27. Prabhakaran, M. and S.C. Harvey, *Molecular-dynamics of structural transitions and intercalation in DNA.* Biopolymers, 1988. **27**: p. 1239-1248.

28. Rao, S.N. and P.A. Kollman, *Molecular Mechanical Simulations on Double Intercalation of 9-Amino Acridine into d(CGCGCGC)-d(GCGCGCG) - Analysis of the Physical Basis for the Neighbor-Exclusion Principle.* Proc Natl Acad Sci U S A, 1987. **84**: p. 5735-5739.

29. Marin V., H.H.F., Koch T., Armitage B.A., *Effect of LNA Modifications on Small Molecule Binding to Nucleic Acids.* J Biomol Struct Dyn, 2004. **21**: p. 841-850.

30. Goddard, N.L., et al., *Sequence Dependent Rigidity of Single Stranded DNA.* Phys Rev Lett, 2000. **85**: p. 2400-2403.

31. Cafferty, B.J. and N.V. Hud, *Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology.* Curr Opin Chem Biol, 2014. **22**: p. 146-157.

32. Engelhart, A.E., M.W. Powner, and J.W. Szostak, *Functional RNAs exhibit tolerance for non-heritable 2 '-5 ' versus 3 '-5 ' backbone heterogeneity.* Nature Chem, 2013. **5**: p. 390-394.

CHAPTER 9

A PCR COMPATIBLE OLIGONUCLEOTIDE LIGATION SYSTEM WITH AN INTERMEDIATE REVERSIBLE STEP[h]

9.1. Introduction

DNA is now used for purposes well beyond its natural information-carrying capacity (e.g., in nanotechnology, as a catalyst, and as a recognition element) [1-5]. As a result, this polymer appears in applications as varied as sensors, therapeutic agents, and imaging technologies [5-7]. Synthetic procedures for the nonenzymatic coupling of oligonucleotides that are compatible with self-assembly in aqueous solution, allowing selection of thermodynamically favored assemblies, have further expanded the possible uses of DNA [8, 9]. However, most methods used for the nonenzymatic polymerization of nucleic acids lack a means to allow thermodynamic selection, involve synthetic procedures that are inaccessible to most laboratories, or result in formation of a polymer linkage that polymerase enzymes cannot read, thereby limiting the accessibility and utility of existing ligation methods. This chapter discusses a DNA ligation system that: a) is accessible via commercially available reagents; b) includes a reversible step that can allow selection of a thermodynamically favored product; c) proceeds in a template-selective mode at lower concentrations; d) proceeds in an untemplated mode at higher substrate concentrations; and e) produces a linkage that can be tolerated in a template

h. This chapter was adapted from previously published work and is reproduced with permission. Engelhart, AE; Cafferty, BJ; Okafor, CD; Chen, MC; Williams, LD; Lynn, DG; Hud, NV "Nonenzymatic ligation of DNA with a reversible step and final linkage that can be used in PCR." *ChemBioChem* (2012).

strand by a number of thermophilic and mesophilic polymerases, allowing product sequence amplification by PCR.

It has been known for some time that the 2′,3′ dialdehydes generated by periodate oxidation of ribonucleotides react in water with alkylamines to produce, upon reduction, a hydrolytically stable morpholine [10]. Wincott and coworkers have previously shown the utility of this amine-dialdehyde ligation reaction within an RNA hairpin loop [11], and other groups have used similar reactions to extend and conjugate chemical groups onto the ends of nucleic acids [11-16]. However, the effect of a morpholine linkage on DNA duplex stability and in a PCR template strand has not been reported. In this study the morpholine backbone linkage is investigated for its compatibility with DNA duplex stability and with various polymerases. The initial, reversible, linkage can be formed by the reaction of an oligonucleotide containing a 5′-amino-dT residue with an oligonucleotide containing a periodate-oxidized 3′-ribose followed by reduction by sodium cyanoborohydride (Figure 9.1). The linkage investigated here differs substantially from the well-known phosphorodiamidate morpholino linkage [17] in that reduced linkage lacks a phosphorodiamidate group between the moropholine ring and the residue on its 3′ side.

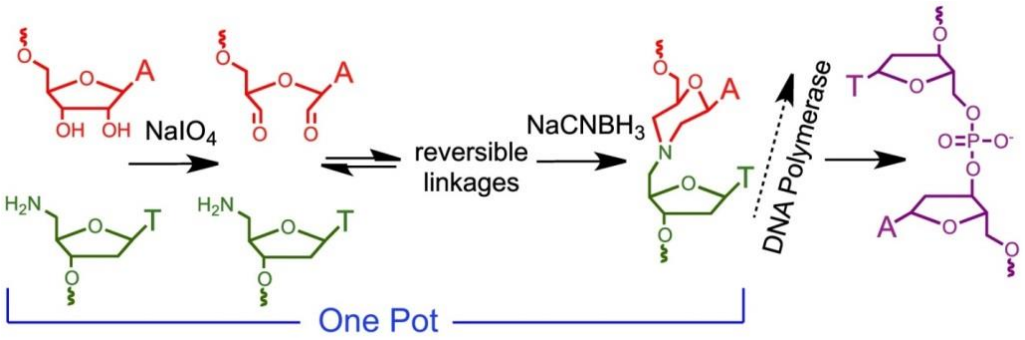


Figure 9.1 Diagram of morpholine ligation reactions and polymerase read-through.

9.2. Experimental Procedures

9.2.1. Materials

Sodium cyanoborohydride and sodium periodate were purchased from VWR. 5′-amino-dT oligonucleotides were prepared using an ABI Expedite 8909 using phosphoramidites, supports, and reagents from Glen Research. These oligonucleotides were prepared MMT-on, deprotected by 48 hr incubation in concentrated ammonium hydroxide at room temperature, purified by RP-HPLC, MMT removed with 20% acetic acid 1 hr at RT, neutralized with triethylamine, and desalted on a Sep-Pak Plus C18 cartridge (Waters). All other oligonucleotides were purchased from IDT, purified by IE-HPLC (Dionex DNAPac PA-100), and desalted on a Sep-Pak Plus C18 cartridge. All oligonucleotides were characterized by ESI-MS. Deep VentR (exo-), Klenow Fragment, Taq, Phi29, Bsu, AMV RT, Bst, Tth, Therminator, 9° North, Vent (exo-) and M-MuLV RT were purchased from New England Biolabs. SuperScript II RT was purchased from Invitrogen. Sequenase was purchased from GE Healthcare. HIV RT was purchased from Worthington Biochemical.

Table 9.1 Nucleic acids used in the morpholine linkage study.

Name	Sequence
Am5	d(H_2N-TAA GC)
Ald5	d(GAG T)rC
Ald5MM	d(GAA T)rC
Temp10	d(GAC TCA TTC G)
Phos10	d(CGA ATG AGT C)
T_{15}	d(TTT TTT TTT TTT TTT)
Am49	d(H_2N-T CTC GAA TGG CTT ACT GAC ACC CTA TAG TGA GTC GTA TTA GAA TTC CGG)
Ald23	d(CGT AGA TCC GCG GAC TAT GAG G)rA
Splint21	d(CCA TTC GAG ATC CTC ATA GTC)
Bottom72	d(CGT AGA TCC GCG GAC TAT GAG GAT CTC GAA TGG CTT ACT GAC ACC CTA TAG TGA GTC GTA TTA GAA TTC CGG)
Pause49	d(CCG GAA TTC TAA TAC GAC TCA CTA TAG GGT GTC AGT AAG CCA TTC GAG A)
Bottom98	d(CAA AAG GGT CAG TGC TGG ATT ATG CTG AGT GAT ATC CCA CAG TCA TTC GGT AAG CTC TAG GAG TAT CAG GCG CCT AGA TGC CAG TAT CGA CAA AGG AC)
Top98	d(CAA AAG GGT CAG TGC TGG ATT ATG CTG AGT GAT ATC CCA CAG TCA TTC GGT AAG CTC TAG GAG TAT CAG GCG CCT AGA TGC CAG TAT CGA CAA AGG AC)
TopPrimerA	d(CCG GAA TTC TAA TAC GAC TCA CTA TAG)
TopPrimerB	d(CTA ATA CGA CTC ACT ATA GGG TGT C)
BottomPrimerA	d(CGT AGA TCC GCG GAC TAT G)
M13F-TopExtender	d(GTT TTC CCA GTC ACG ACC TAA TAC GAC TCA CTA TAG GGT GTC)
M13R-BottomExtender	d(CAG GAA ACA GCT ATG ACC GTA GAT CCG CGG ACT ATG)
M13F	d(GTT TTC CCA GTC ACG AC)
M13R	d(CAG GAA ACA GCT ATG AC)

9.2.2. Thermal Denaturation Analysis and T_m Determination

Reactions were prepared with 1 mM each of Temp10, Ald5, and Am5. These solutions were held at 0 °C overnight with 2.5 mM $NaIO_4$ to form the reversible linkage (**2**), and with $NaCNBH_3$ added to 25 mM when the morpholine linkage (**1**) was desired. Immediately before a thermal denaturation experiment, an aliquot of the desired stock solution was diluted to 2 µM in each oligonucleotide strand in 100 mM sodium borate, pH 8.5, 25 mM NaCl. For thermal denaturation experiments with the all phosphate-linked DNA backbone, a sample of Phos10 and Temp10 was prepared at 2 µM in each strand, with the same buffer conditions. Each sample was subjected to two cycles of heating and cooling between 5 °C and 50 °C, with spectra taken at 1 °C intervals using an Agilent 8453 UV-vis spectrophotometer equipped with an 89090A temperature controller. The resulting spectra were analyzed by monitoring absorbance at 260 nm (baseline corrected by subtracting absorbance at 350 nm). The T_m for each sample was determined as the midpoint of a sigmoid fit of the resulting transition.

9.2.3 Measurement of Morpholine Reduction Kinetics

Reactions contained 10 µM of each Am5, Ald5 (or Ald5MM for mismatch studies), and Temp10 (or d(T15) as an internal standard in nontemplated reactions), and were in 100 mM sodium borate buffer, pH 8.5. Oligonucleotides were allowed to react with 2.5 mM NaIO4 for 30 min on ice before the addition of 25 mM $NaCNBH_3$ and incubation on ice. Reactions were analyzed by IE-HPLC (Dionex DNAPac PA-100) with a gradient of 1 to 100% 2 M NaOAc, pH 7. Competition reactions carried out at three different concentrations, but with the same relative strand concentrations. For the 10 µM

reactions, oligonucleotide concentrations were Am5, 10 μM; Ald5, 20 μM; Ald5MM, 20 μM; Temp10 or T15, 10 μM. For the 5 μM reactions: Am5, 5μM; Ald5, 10 μM; Ald5MM 10 μM; Temp10 or T15, 5 μM. For the 1 μM reactions: Am5, 1 μM; Ald5, 2 μM; Ald5MM 2 μM; Temp10 or T15, 1 μM. Analysis was performed as above with pH 12 NaOAc mobile phase.

9.2.4 Stoichiometric Readout Reactions

The Ald23-1-Am49 template was generated in a reaction containing 1 mM each Am49 and Ald23, and, optionally, 1 mM Splint21. The DNA construct was purified by separation from substrates and Splint21 (if present) using denaturing PAGE and extracted from the gel using the crush-and-soak method. Bottom72 was used as a template, instead of Ald23-1-Am49, when the all phosphate-linked template was needed for the control reactions. Reactions contained 250 nM in template (Bottom72 or Ald23-**1**-Am49) and 50 nM 32P-labeled TopPrimerA, prepared using T4 PNK (New England Biolabs). Reactions were 10 μL and contained 0.25 μL of the polymerase as supplied by the manufacturer in the supplied buffer at 1X concentration. For reactions containing mesophilic polymerases, the primer and template strands were heated to 95 °C and slow annealed before addition of the polymerase, and were then incubated at 37 °C for 1 h. Reactions containing thermophilic polymerases were incubated at 95 °C for 5 min, then 55 °C for 1 h. After incubation, reactions were diluted with 10 μL 8 M urea, 1X TBE, separated on a 10% denaturing polyacrylamide gel, and analyzed using a phosphorimager (GE Typhoon). Quantification of full-length read-through is reported as the integrated

intensity of the 72 nt band divided by the total integrated intensity of the 72 nt band and the abortive product band(s) at or near 49 nt.

9.2.5 PCR Reactions

Reactions contained 1 nM template (Ald23-1-Am49 or Phos72), 2 μM each of TopPrimerB and BottomPrimerA, 200μM dNTPs, 2X ThermoPol buffer. Reactions were subjected to 4 min at 95 °C, 30 cycles of 2-step PCR (30 sec at 95 °C, 60 sec at 72 °C), and a 5 min final extension at 72 °C. Products were separated from remaining primers on a 10% denaturing polyacrylamide gel, stained with SYBR Gold, and visualized with a UV transilluminator/camera.

9.2.6 M13 Primer Extension and Sequencing

The reaction mixture from the above procedure was diluted to 1 nM (given the assumption that primers were fully converted to full-length product) and amplified with Phusion high fidelity polymerase, 1 nM each M13F-TopExtender (a M13F-TopPrimerB conjugate) and M13R-BottomExtender (a M13R-BottomPrimerA conjugate), and 1 μM M13F and M13R, yielding the 98bp PCR product Top98-Bottom98, which then was gel purified and sequenced (Genewiz) using the M13F primer.

9.2.7 MD Analysis

Molecular models of duplexes Ald23-**1**-Am49·Temp10 and Phos10·Temp10: Molecular dynamics (MD) simulations were performed using the AMBER11 suite of programs with the parmbsc0 modifications to the parm99 force field.

9.3 Results and Discussion

9.3.1 Stability of DNA Duplexes Containing Reversible and Irreversible Linkages

To explore the potential for a DNA oligonucleotide containing a morpholine linkage (**1**) or the related unreduced linkages (including the dihydroxymorpholine linkage, **2**) to support duplex formation, the pentamer d(GAGT)rC (Ald5) was incubated with $NaIO_4$ (to generate the 2', 3'-dialdehyde) in the presence of 5'-amino-d(TAAGC) (Am5) and the decamer d(GCTTAGACTC) (Temp10), which could serve as a template for the ligation of Ald5 and Am5 (Figure 9.2). These oligonucleotide lengths were chosen such that the majority of Ald5 and Am5 would hybridize with Temp10 under the conditions of incubation (1 mm in each strand, 5 °C) [18], with the hybridized complex being further stabilized if either linkage **1** or **2** is compatible with a DNA duplex.

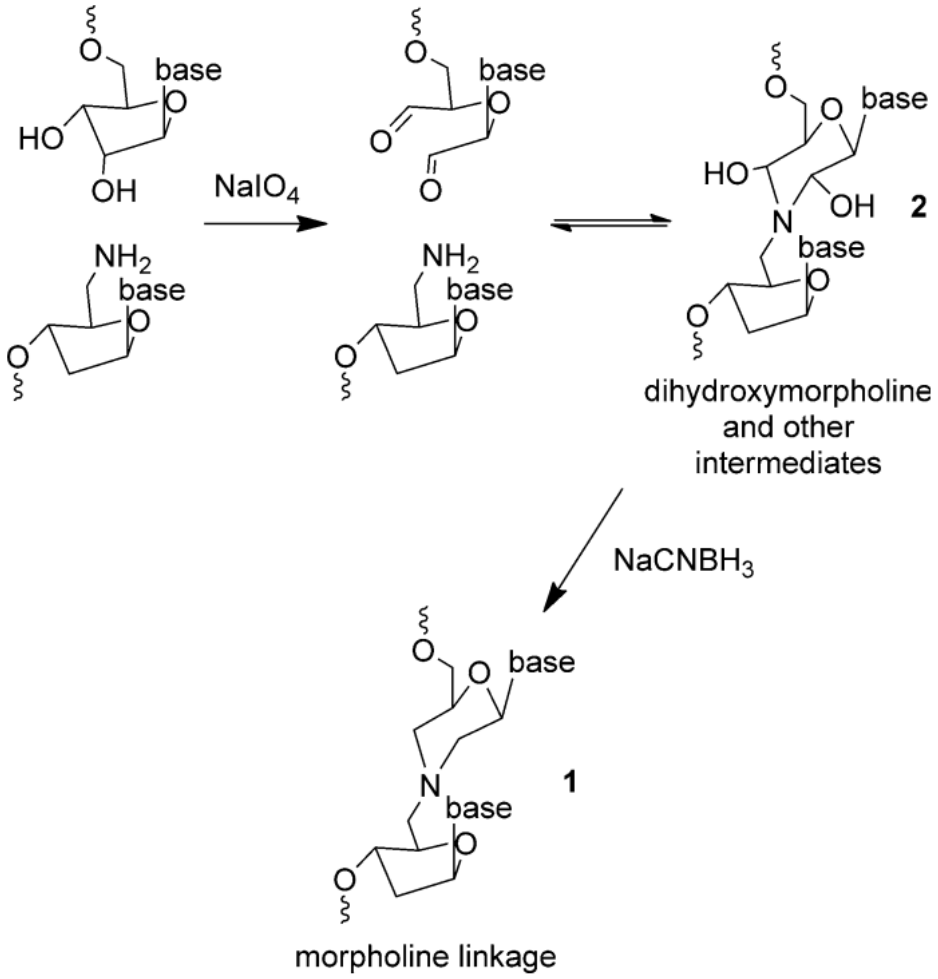


Figure 9.2 Schematic showing each step of the ligation reaction of 3'-ribo-oligonucleotides and 5'-amino-oligonucleotides resulting in morpholine ring formation. **1**, indicates the irreversible linkage and **2**, indicates the reversible linkage.

The stabilizing effect of linkage **2** is demonstrated by the fact that, in the absence of a reducing agent, Ald5, Am5, andTemp10 form a complex, which, when diluted 500-fold immediately before analysis, exhibits a cooperative melting transition (T_m) of less than 20 °C (see Figure 9.3). The T_m of less than 20 °C is reported as a lower bound, because the complex did not reassemble after heating and cooling back to 5 °C. It should be noted that the individual melting temperatures for unlinked Ald5 and Am5 with Temp10 are predicted to be less than 0 °C in the conditions of the diluted sample [18]. Thus, our observation of a melting transition at 20 °C is fully consistent with the dissociation of a reversibly linked Ald5-**2**-Am5 strand from the Temp10 template, followed by the hydrolysis of Ald5-**2**-Am5 back to unlinked Ald5 and Am5.

To examine if the *irreversible* linkage **1**, formed upon the reduction of linkage **2**, is compatible with a DNA duplex, the reducing agent $NaCNBH_3$ was added to a solution containing Ald5, Am5, and Temp10, at a temperature and oligonucleotide concentrations (0 °C; 1 mM in each strand) for which the formation of Ald5-**2**-Am5 had already been verified by melting studies. Following sufficient time for linkage reduction (i.e., 24 h) and dilution to a final concentration of 2 μM in each strand, thermal denaturation of this sample revealed a cooperative transition at 24 °C, which is only 2 °C less than the T_m observed for the duplex formed by Temp10 with a complementary strand containing all phosphodiester linkages. Unlike the pre-reduction duplex, the hydrolytically stable Ald5-**1**-Am5 showed duplex reformation with Temp10 during sample cooling (Figure 9.3). Additionally, molecular dynamics (MD) simulations were carried out on the Ald5-**1**-Am5/Temp10 duplex and suggested that linkage **1** can be maintained in a B-form duplex, but may potentially assume alternative conformations as well [19, 20]. Taken together,

these observations demonstrate that a single substitution of the reduced morpholine linkage is well tolerated within a DNA duplex.

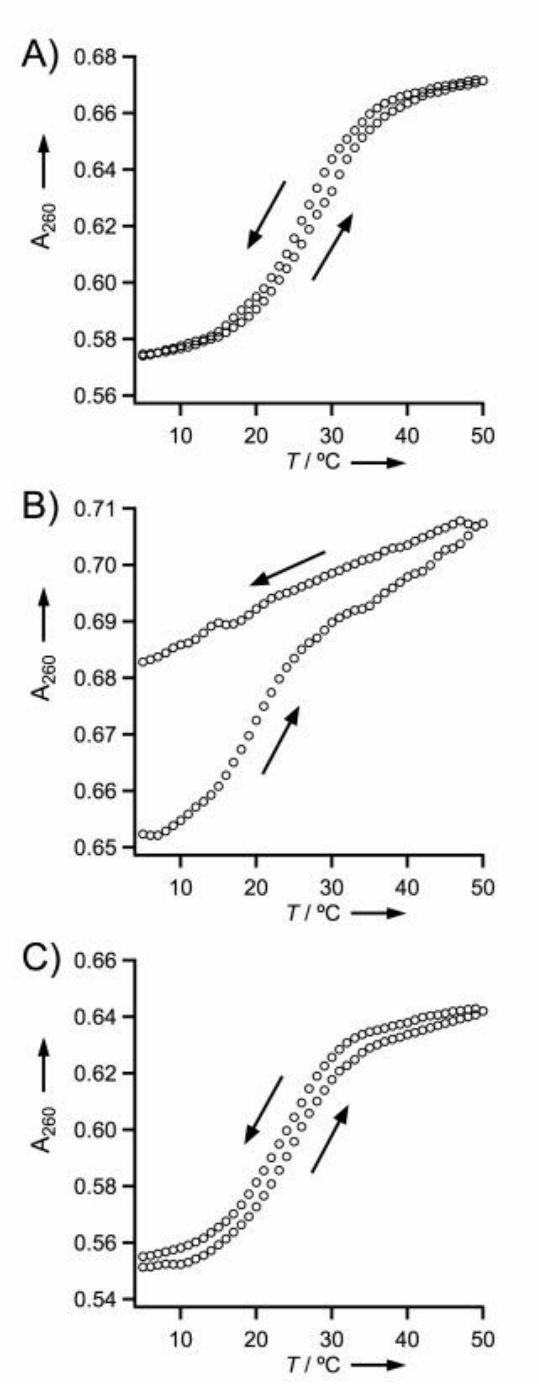


Figure 9.3 Thermal denaturation/renaturation profiles of duplexes with and without single modified linkages. (A) Duplex consisting of Phos10, the all phosphate-linked strand, and its complementary strand (also all phosphate-linked) Temp10. (B) Duplex consisting of Ald5-**2**-Am5, the strand with an unreduced (reversible) linkage, and Temp10. (C) Duplex consisting of Ald5-**1**-Am5, the strand with the reduced (irreversible) morpholine-linkage, and Temp10.

9.3.2 Sequence Specificity of Linkage Formation

The sequence specificity of this template-directed ligation reaction was next examined. The addition of a stoichiometric amount of Temp10 to reaction mixtures containing 10 μM Am5 and Ald5 resulted in a 20-fold enhancement in the rate of Am5-**1**-Ald5 formation relative to the rate of spontaneous Am5-**1**-Ald5 formation that is observed at substrate concentrations of 10 μM and higher (Figure 9.4A). When Ald5 was replaced in the reaction with d(GAAT)rC (Ald5MM), a strand that contains a single mismatch when bound to Temp10, the addition of Temp10 resulted in a Am5MM-**1**-Ald5 formation rate essentially identical to that of a control reaction lacking a template strand (Figure 9.4A). Additionally, in a competition experiment containing both Ald5 and Ald5MM in the presence of Temp10, Ald5 ligated with Am5 with approximately 7-fold higher yield than Ald5MM after two days of reaction (Figure 9.4B). When the competition experiment was performed without Temp10, a similar rate was observed for Am5 ligation with Ald5 and Ald5MM (i.e., the rate of untemplated substrate coupling at 10 μM substrate concentration) (Figure 9.4B). When these reactions were carried out with Am5, Ald5, and Ald5MM at 1 μM in the absence of Temp10, no ligation products were detected after two days.

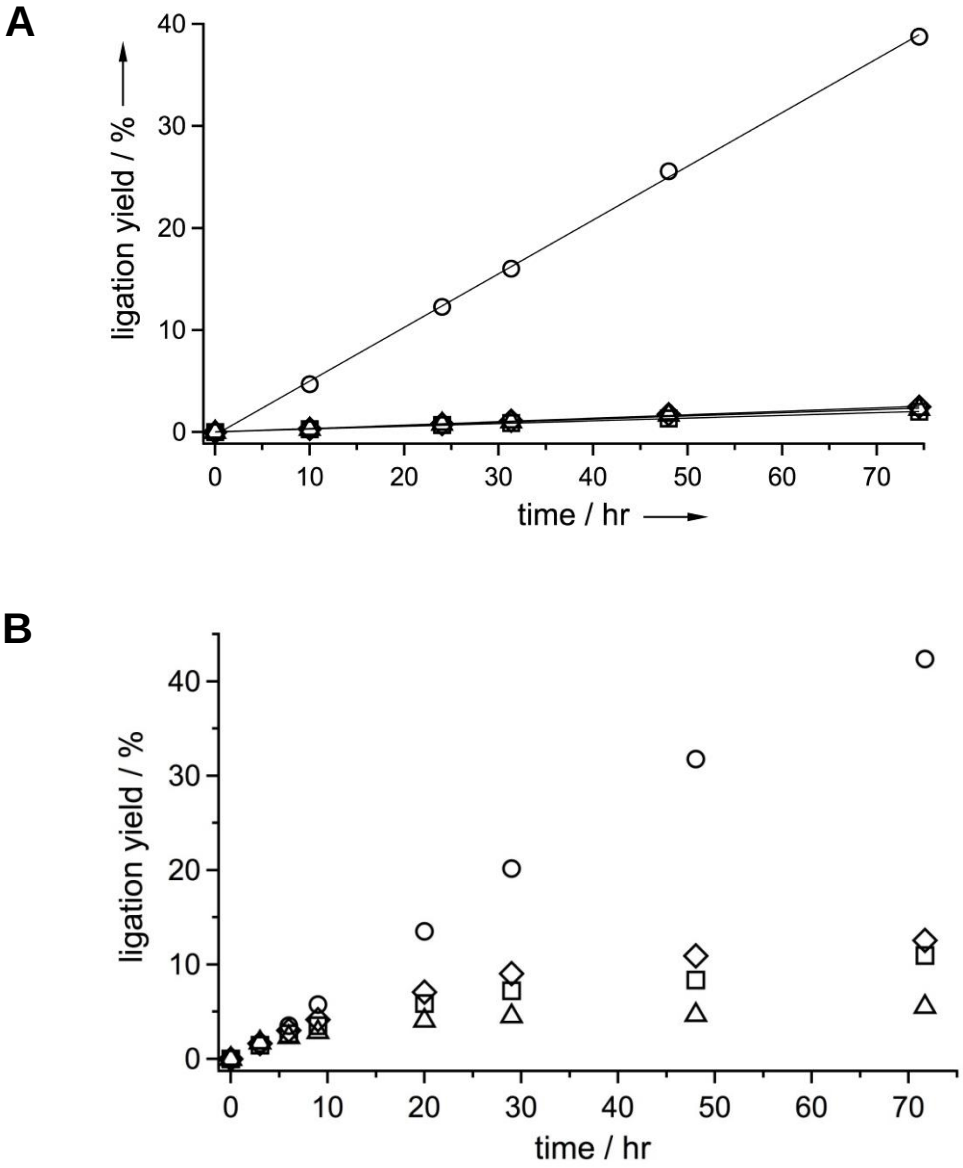


Figure 9.4 Comparison of morpholine-linked product formation rates in templated, and untemplated reactions, with or without an internal mismatch in the dialdehyde substrate. (A) Ald5-**1**-Am5 with Temp10 present (○), Ald5MM-**1**-Am5 with Temp10 present (△), Ald5-**1**-Am5 untemplated (◇); Ald5MM-**1**-Am5 untemplated. (B) Effect of template Temp10 on Ald5-**1**-Am5 and Ald5MM-**1**-Am5 formation (reduction) rates in competition reactions containing both Ald5 and Ald5MM at 10 μM. Ald5-**1**-Am5 with Temp10 present (○); Ald5MM-**1**-Am5 with Temp10 present (△); Ald5-**1**-Am5 untemplated (◇); Ald5MM-**1**-Am5 untemplated (□). Reactions were at 0 °C, with oligonucleotides at 10 μM in strand. Oligonucleotides were reacted with 2.5 mM $NaIO_4$ for 30 min before the addition of 25 mM $NaCNBH_3$.

9.3.3 Primer Extension

Having verified sequence selectivity by this ligation system in a template-directed reaction, the product of a ligation reaction was next evaluated for its ability to function as a template for a naturally occurring DNA polymerase. For these experiments, a ribose-terminated 23-mer (Ald23) and a 5'-amino 49-mer (Am49) were prepared, along with oligonucleotides that could serve to hybridize to Ald23 and Am49, bringing the ribose and amino termini in close proximity. Using these oligonucleotides and the chemical steps described above for aldehyde generation and linkage reduction, Ald23-**1**-Am49 was generated. A number of thermophilic and mesophilic polymerases proved capable of extending a radiolabeled primer when annealed to the morpholine-containing Ald23-**1**-Am49 (Figure 9.5). This result was obtained irrespective of whether Ald23-**1**-Am49 was generated in a templated reaction, using a full-length template (Temp72) or a shorter template (Splint21) that formed a 21-mer duplex between Am49 and Ald23, or by the spontaneous higher-concentration ligation reaction of Am49 and Ald23 (i.e., in the absence of a template). When a mixture of Ald23, Am49, and Splint21 was employed without periodate treatment (i.e., without dialdehyde generation), no 72-mer product was observed, indicating that overlap extension was not providing a false positive by isothermal overlap extension (Figure 9.6).

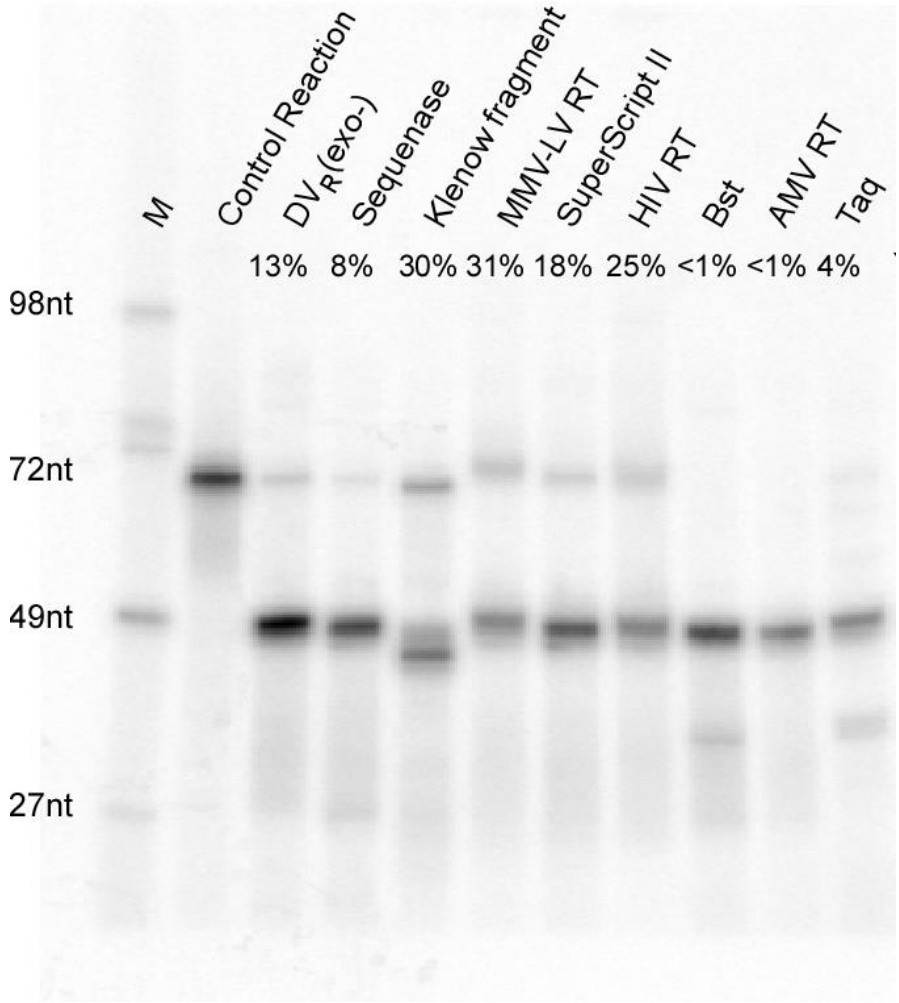


Figure 9.5 Autoradiograph of PAGE-separated products resulting from primer extension by selected enzymes with Ald23-**1**-Am49 as a template. Yields of the full-length 72-mer product for each enzyme are given after enzyme names. Control Reaction is product of the all phosphate-linked template by DVR (exo-).

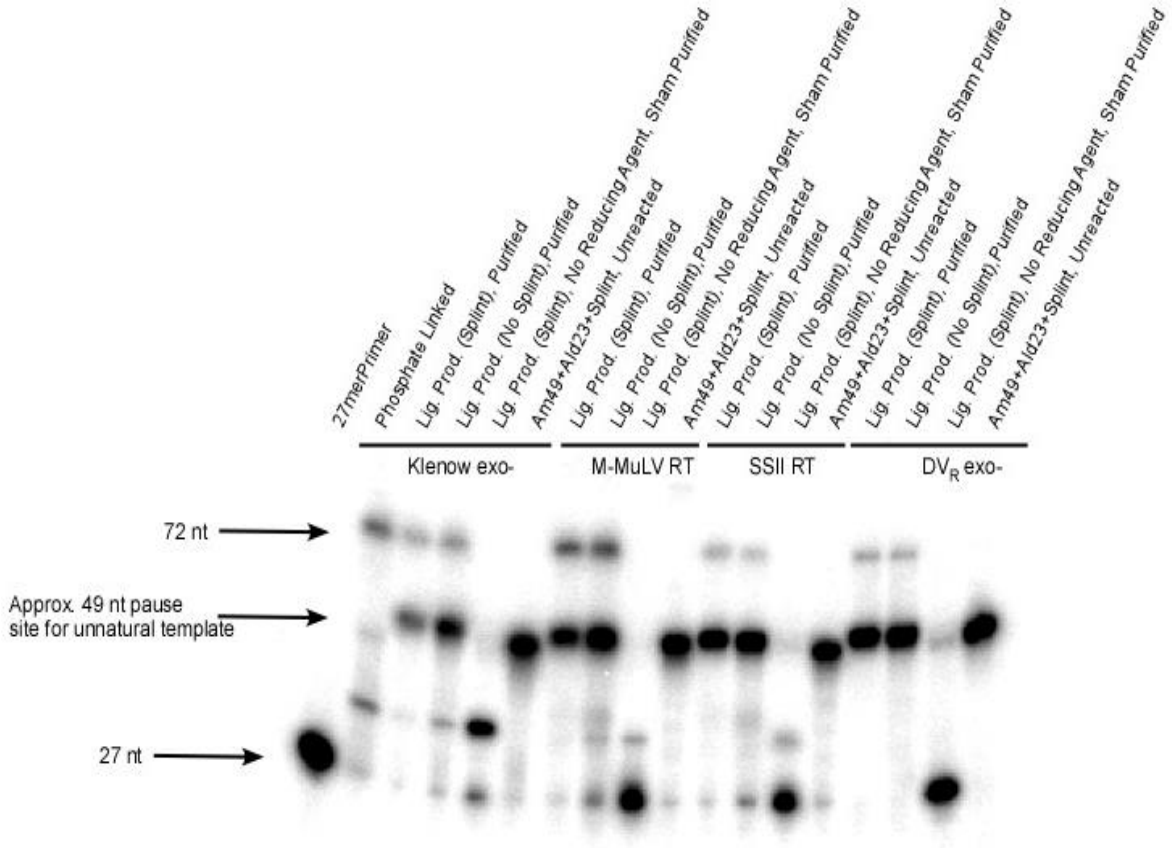


Figure 9.6 Verification by PAGE analysis that the morpholine linkage is required for full-length product in stoichiometric primer extension reactions. Several polymerases produced full-length product when the PAGE-purified Ald23-**1**-Am49 ligation product (Lig. Prod.) was supplied as a template, regardless of whether Ald23-**1**-Am49 was produced in the presence of Splint21 (Splint) strand or by the untemplated ligation (No Splint) of Ald23 and Am49. To verify that full-length product was not the result of overlap extension, reactions containing the ligation substrates and the split strand, but without oxidation of the Ald23 terminal ribose. The absence of full-length product for the unreacted substrates demonstrates that Ald23-**1**-Am49 is required for full-length product and that overlap extension is not appreciable in these experiments.

For all primer-extension reactions with Ald23-**1**-Am49 as a template, the primer was not fully extended in 100% yield for any polymerase. A significant amount of 49 nt product was generated, consistent with all of the polymerases tested pausing, or even stopping, at or near the site of the morpholine linkage (Figure 9.5 and 9.6). Quantification of full-length product yield revealed a range of propensity to read through linkage **1**, from no detectable full-length product (Bst, AMV RT) to more than 25% (Klenow fragment, MMLV RT, HIV RT) (see lane labels in Figure 9.5). The production of an appreciable fraction of abortive products is not necessarily problematic, however, given that, in most cases, the fraction of primer fully extended was at least 4%, and PCR could be used as a subsequent reaction to amplify the full-length product.

9.3.4 PCR Amplification of a DNA Construct Containing a Morpholine Linkage

One enzyme that produced considerable full-length product with the morpholine-containing Ald23-**1**-Am49, Deep VentR (exo-) DNA polymerase (DVR (exo-), New England Biolabs), is a high-fidelity thermophilic polymerase routinely used in PCR. Using DVR (exo-) with Ald23-**1**-Am49 and primers designed to yield a 64 bp product, a PCR reaction resulted in a one-pot readout-amplification reaction, as indicated by the observation of a product of expected length and a gel mobility identical to the PCR product of an all-phosphate-linked template of identical sequence to Ald23-**1**-Am49 (Figure 9.7).

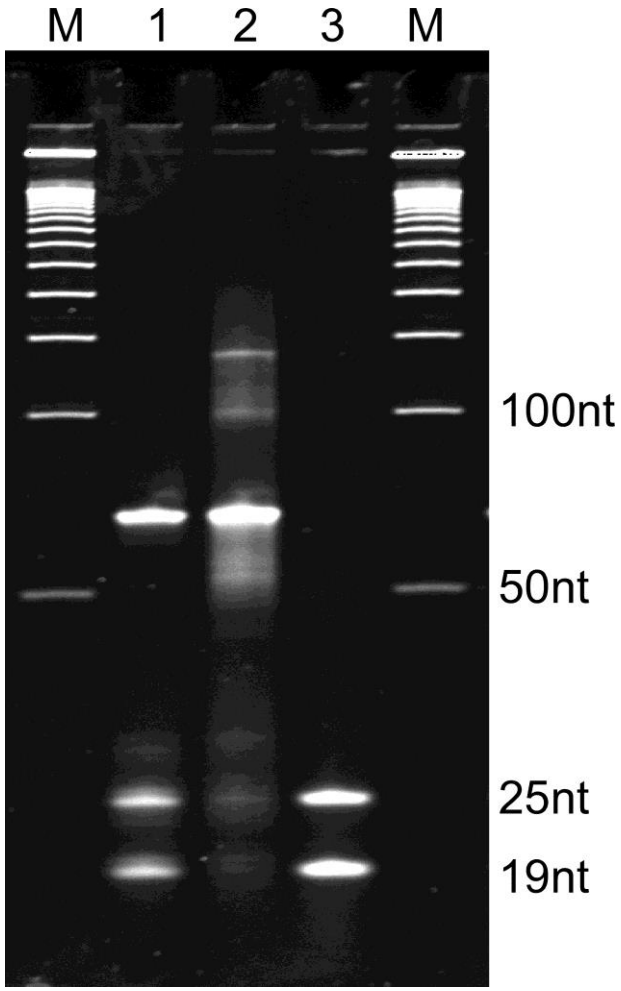


Figure 9.7. PAGE analysis of DVR (exo-) amplification of a sequence containing a morpholine linkage. Ald23-**1**-Am49 (Lane 1), an all phosphate-liked template with a sequence identical to Ald23-**1**-Am49 (Lane 2), and control reaction with PCR primers and no template (Lane 3). Marker bands (Lanes M) are 50nt ladders. 25 and 15nt bands represent PCR primers. Reactions contained 1 nM template, 2 µM of each primer, 200 µM dNTPs, 2X ThermoPol buffer. Thermal cycle: 4 min at 95 °C, 30 cycles of 2-step PCR (30 sec at 95 °C, 60 sec at 72 °C), and a 5 min final extension at 72 °C.

To investigate the fidelity of template read-through across the morpholine linkage, the 64 bp PCR product described above was subjected to a second round of PCR (using Phusion high-fidelity polymerase, New England Biolabs) that added the M13 17-mer forward and reverse primer sites for sequencing. The resulting 98 bp product was PAGE-purified and sequenced. The isolated dsDNA retained the initial Ald23-**1**-Am49-derived template sequence with no mutations, including at or near the morpholine linkage site (Figure 9.8). This result was obtained regardless of whether the original Ald23-**1**-Am49 template strand was produced by template-directed ligation with Splint21 or by spontaneous ligation, illustrating that both the untemplated and templated reaction are capable of generating templates suitable for PCR amplification.

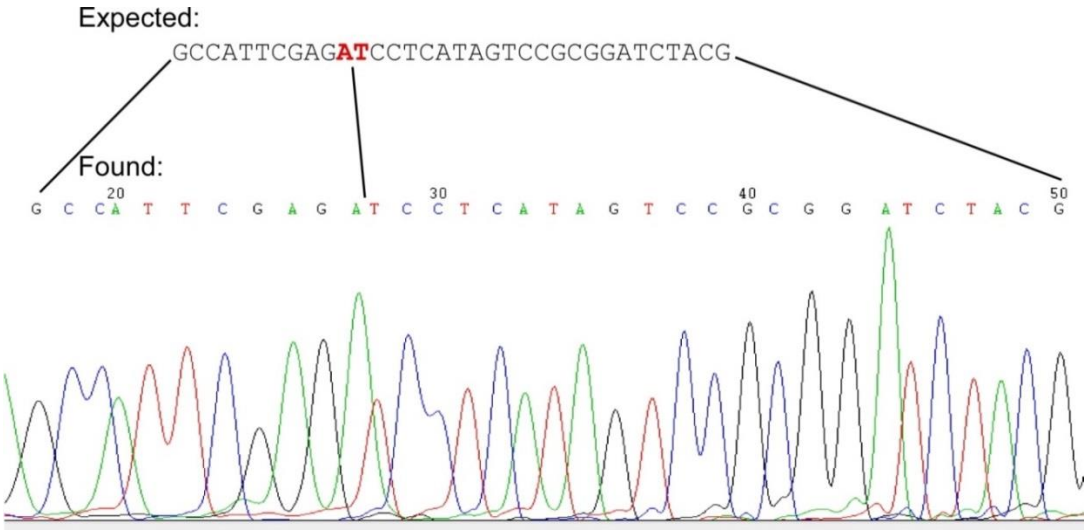


Figure 9.8 Sequencing results of PCR product originating from a Ald23-**1**-Am49 template. Sequencing data obtained with M13F and M13R primers for the Top98-Bottom98 duplex that was produced by Deep $Vent_R$ (exo-) read-through/amplification of Ald23-**1**-Am49 and subsequent addition of the M13 primer sites using Phusion high fidelity polymerase, as described in the PCR procedure. The sequencing read corresponding to the region at and around the morpholine linkage in Ald23-**1**-Am49 is shown. The data shown is for a PCR product produced using Ald23-**1**-Am49 template that was produced in a templated ligation reaction with Splint21. Sequencing data for PCR product using Ald23-**1**-Am49 from an untemplated ligation reaction yielded the same sequence, and virtually identical peak heights.

9.4 Conclusions

Towards finding nonenzymatic aqueous DNA ligation chemistries that form linkages capable of being utilized by polymerases, Brown and coworkers recently demonstrated that a modified oligonucleotide containing a single neutral triazole linkage, a product of "click" Huisgen cycloaddition [23], can be amplified by PCR and replicated *in vivo* [24-26]. In their development of ligation chemistries, Brown and coworkers observed that their first generation linkage suffered from nucleotide deletion at the linkage site after PCR amplification, a result that was interpreted as being symptomatic of a linkage that was too rigid [27]. This body of work by Brown and coworkers illustrated the importance of a modification's ability to structurally mimic that of the natural phosphodiester linkage. As suggested by our modeling studies, while linkage **1** does differ markedly from the canonical backbone linkage, it appears to, like the second-generation linkage of Brown and coworkers, be able to adopt a structure that is acceptable for use by a polymerase. After the original publication of the work described herein, Gothelf and coworkers published on a reversible disulfide linkage that could also be 'read-through' but polymerases and used in PCR [28]. Upon sequencing, the authors also noted a relatively high rate of mutation at the site of the disulfide linkage.

Finally, the polymerases used in our single primer extension studies showed various abilities to tolerate a morpholine linkage in a template strand (Figure 9.5). There are correlations between our polymerase-screening study and those that have been previously reported for the reading of templates that contain other unnatural linkages. For example, Szostak and coworkers reported that MMLV RT, Sequenase and SSII RT were among the polymerases with the greatest ability to read templates with sequential

residues of threose-nucleic acid, or TNA, whereas the polymerase AMV RT was one of the least able to read TNA [29]. The same grouping of these polymerases are observed for their ability to read through a morpholine linkage. In contrast, the polymerase Bst, which was also reported to read a TNA template, did not show appreciable read-through of the morpholine linkage. In another study that utilized templates containing 2′,5′-linked DNA, Switzer and coworkers reported that HIV RT and Klenow (exo-) worked better than AMV RT and Taq [30]. These results are also consistent with the ranking of these enzymes for toleration of a morpholine linkage. Taken together, these comparisons of polymerase activity with templates containing non-natural linkages provide additional evidence that some polymerases are intrinsically more accepting of non-natural linkages than others, and that some polymerases may be very selective regarding which unnatural linkages will be tolerated.

In conclusion, this work has demonstrated that a morpholine backbone linkage, introduced into an otherwise natural DNA strand, can be tolerated as a linkage in the template strand of a PCR reaction with a naturally-occurring polymerase. This linkage can be generated in a one-pot reaction, using commercially available reagents, enhancing the general accessibility of this chemistry. Several features of the morpholine linkage system discussed here promise to make this chemistry, and related systems, useful for a wide range of applications. In particular, the existence of a covalent linkage, which is formed in water and is reversible until reduced, provides the ability to select thermodynamically favored products during nucleic acid self-assembly [8, 9]. Additionally, the enzyme-free nature of this ligation system has the potential to be applied in non-duplex regions of nucleic acids (e.g., in loops, triplexes, and G

quadruplexes), such as during dynamic combinatorial reactions, followed by post-reduction read-through by a polymerase.

9.5 References

1. Seeman, N.C., *From genes to machines: DNA nanomechanical devices.* Trends Biochem Sci, 2005. **30**: p. 119-125.

2. Aldaye, F.A., A.L. Palmer, and H.F. Sleiman, *Assembling materials with DNA as the guide.* Science, 2008. **321**: p. 1795-1799.

3. Silverman, S.K., *DNA as a Versatile Chemical Component for Catalysis, Encoding, and Stereocontrol.* Angew Chem Int Ed, 2010. **49**: p. 7180-7201.

4. Poon, L.C.H., et al., *Guanine-Rich RNAs and DNAs That Bind Heme Robustly Catalyze Oxygen Transfer Reactions.* J Am Chem Soc, 2011. **133**: p. 1877-1884.

5. Cho, E.J., J.W. Lee, and A.D. Ellington, *Applications of Aptamers as Sensors.* Annu Rev Anal Chem, 2009. **2**: p. 241-264.

6. Ng, E.W.M., et al., *Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease.* Nat Rev Drug Discov, 2006. **5**: p. 123-132.

7. Stojanovic, M.N. and D.M. Kolpashchikov, *Modular aptameric sensors.* J Am Ch Soc, 2004. **126**: p. 9266-9270.

8. Li, X., et al., *Step-growth control in template-directed polymerization.* Heterocycles, 2011. **82**: p. 1477-1488.

9. Zhan, Z.Y.J. and D.G. Lynn, *Chemical amplification through template-directed synthesis.* J Am Chem Soc, 1997. **119**: p. 12420-12421.

10. Ermolinsky, B.S. and S.N. Mikhailov, *Periodate oxidation in chemistry of nucleic acids: Dialdehyde derivatives of nucleosides, nucleotides, and oligonucleotides (review).* Bioorganicheskaya Khimiya, 2000. **26**: p. 483-504.

11. Bellon, L., et al., *Morpholino-linked ribozymes: A convergent synthetic approach.* J Am Chem Soc, 1996. **118**: p. 3771-3772.

12. Wang, Z.F., et al., *Convergent synthesis of ribonuclease L-active 2 ',5 '-oligoadenylate-peptide nucleic acids.* Bioorg Med Chem Lett, 2000. **10**: p. 1357-1360.

13. Zhou, L.H., et al., *Endowing RNase H-inactive antisense with catalytic activity: 2-5A-morphants.* Bioconjugate Chem, 2005. **16**: p. 383-390.

14. Kojima, N., et al., *Enhanced reactivity of amino-modified oligonucleotides by insertion of aromatic residue.* Bioorg Med Chem Lett, 2006. **16**: p. 5118-5121.

15. Bell, N.M., R. Wong, and J. Micklefield, *A Non-Enzymatic, DNA Template-Directed Morpholino Primer Extension Approach.* Chem Eur J, 2010. **16**: p. 2026-2030.

16. Wong, O.Y., P.I. Pradeepkumar, and S.K. Silverman, *DNA-Catalyzed Covalent Modification of Amino Acid Side Chains in Tethered and Free Peptide Substrates.* Biochemistry, 2011. **50**: p. 4741-4749.

17. Summerton, J. and D. Weller, *Morpholino antisense oligomers: design, preparation, and properties.* Antisense Nucleic Acid Drug Dev., 1997. **7**: p. 187-195.

18. SantaLucia, J. and D. Hicks, *The thermodynamics of DNA structural motifs.* Annual Review of Biophysics and Biomolecular Structure, 2004. **33**: p. 415-440.

19. Case, D.A., et al., *Amber 11*. 2010, University of California: San Francisco.

20. Dupradeau, F.-Y., et al., *The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building.* Phys. Chem. Chem. Phys., 2010. 12: p. 7821-7839.

21. Hall, H.K., Jr., *Correlation of the Base Strengths of Amines.* J Am Chem Soc, 1957. **79**: p. 5441-5444.

22. Lavery, R., et al., Nucleic Acids Res., 2009. **37**: p. 5917–5929.

23. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click chemistry: diverse chemical function from a few good reactions.* Angew Chem Int Edit, 2001. **40**: p. 2004.

24. El-Sagheer, A.H., et al., *Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli.* Proc Natl Acad Sci U S A, 2011. **108**: p. 11338-11343.

25. Birts, C.N., et al., *Transcription of click-linked DNA in human cells.* Angew Chem Int Ed Engl, 2014. **53**: p. 2362-2365.

26. Sanzone, A.P., et al., *Assessing the biocompatibility of click-linked DNA in Escherichia coli.* Nucleic Acids Res, 2012. **40**: p. 10567-10575.

27. El-Sagheer, A.H. and T. Brown, *Synthesis and Polymerase Chain Reaction Amplification of DNA Strands Containing an Unnatural Triazole Linkage.* J Am Chem Soc, 2009. **131**: p. 3958-3964.

28. Hansen, D.J., et al., *Synthesis, dynamic combinatorial chemistry, and PCR amplification of 3'-5' and 3'-6' disulfide-linked oligonucleotides.* Angew Chem Int Ed Engl, 2014. **53**: p. 14415-14418.

29. Chaput, J.C. and J.W. Szostak, *TNA synthesis by DNA polymerases.* J Am Chem Soc, 2003. **125**: p. 9274-9275.

30. Sinha, S., P.H. Kim, and C. Switzer, *2',5'-linked DNA is a template for polymerase-directed DNA synthesis.* J Am Chem Soc, 2004. **126**: p. 40-41.

CHAPTER 10

Conclusion and Future Directions[f]

We not only want to know how nature is, but we also want to reach, if possible, a goal which may seem utopian and presumptuous, namely, to know why nature is such and not otherwise.

Albert Einstein (1929)

10.1 From Molecular Self-Assembly to Proto-RNA

The water-based synthesis of 2,4,6-triaminopyrimidine (TAP, for structure see **58** in Figure 1.5), melamine (MA, **85**) and barbituric acid barbituric acid (BA, **79**) nucleosides and nucleotides, and their ability to assemble into linear fibers, may represent a significant advance in our understanding of how matter can self-organize into structures that gave rise to proto-RNA. Such supramolecular assemblies may have selected pairing bases out of complex environments, through Watson-Crick-like interactions, and organized functional groups to promote polymerization reactions. However, non-covalent structures such as stacked hexads *do not* contain evolvable information, which would only emerge upon the covalent linking of monomers (i.e., polymerization). Below two models are discussed by which information polymer may have emerged from supramolecular polymers.

f. This chapter was adapted from previously published work and is reproduced with permission. Cafferty, BJ; Hud, NV. "Was a pyrimidine-pyrimidine base pair the ancestor of Watson-Crick base pairs? Insights from a systematic approach to the origin of RNA" Isr. J. Chem. (2015). Cafferty, BJ; Hud, NV. "Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology" Chem. Biol. (2014). Hud, NV; Cafferty, B.J.; Krishnamurthy, R; Williams, LD. "The origin of RNA and "My gradfathers axe" Chem. Biol. (2013).

The Self-Sorting Supramolecular Polymer Model

It is not obvious how a particular sugar would have initially been selected for use in the proto-RNA backbone, a problem compounded by non-enzymatic replication (e.g., enantiomeric cross-inhibition), however the self-sorting behavior of supramolecular assemblies may provide a solution. The noncovalent assembly of a given set of monomers containing the same recognition elements (e.g., TAP, BA, MA or Cy) but different exocyclic substitutions (e.g., different sugars) into supramolecular polymers would provide a mechanism by which backbone structure could be selected, and even limited amounts of functional monomer collected, prior to coupling [1]. Supramolecular polymers are dynamic, with a constant exchange of monomers from solution with monomers of the assemblies (as was demonstrated in Chapter 3 for the TAP-CyCo6-CyCo4 assemblies); while also being under equilibrium control, meaning that the most thermodynamically favored structures will persist and therefore accumulate. For this model, self-assembling heterocycles will react to form monomers with different substitutions, some of which will be more prone to assemble and/or mutually react. Through self-sorting, the monomers will experience numerous arrangements within the assemblies, and will eventually present monomers in neighboring positions that possess compatible geometries and functionalities to form linkages (i.e., backbone connections). Overtime, covalent polymers containing information will be synthesized. Many examples of self-sorting among supramolecular assemblies exist, demonstrating that monomer selection by size/shape discrimination, complementarity and even chirality (from a solution of racemic monomers) can be accomplished [2, 3].

The Polymer Fusion Model

We previously proposed a model for the formation of proto-RNA through the fusion of two polymer types, one a supramolecular polymer (e.g., rosette assemblies) and the other a covalently linked polymer (e.g., polysaccharides or polyesters) (Figure 10.1) [4]. Briefly, in a solution of coexisting supramolecular polymers and a variety of covalently linked polymers, proto-RNA could be formed if there was a specific complementarity between the functional groups on both polymer types. That is, functional groups attached to assembled recognition units, on the solvent exposed periphery of the supramolecular polymer, would match both the spacing and reactivity of the chemical groups along one or a subset of covalently linked polymer(s). Once coupled together, a single covalent polymer capable of storing information would result. We note that the self-sorting supramolecular polymer model and the polymer fusion model are not mutually exclusive, as the former model provides a means in which short covalently linked RNA-like polymers can be moved within and between noncovalent assemblies to find and link with other oligomers. Additionally, since both models require the formation of thermodynamically accessible (i.e., hydrolytically reversible) bonds between the heterocycle and the trifunctional linkers, these theories use dynamic combinatorial chemistry-based approaches, which have been discussed in the context of the origin of life [5, 6].

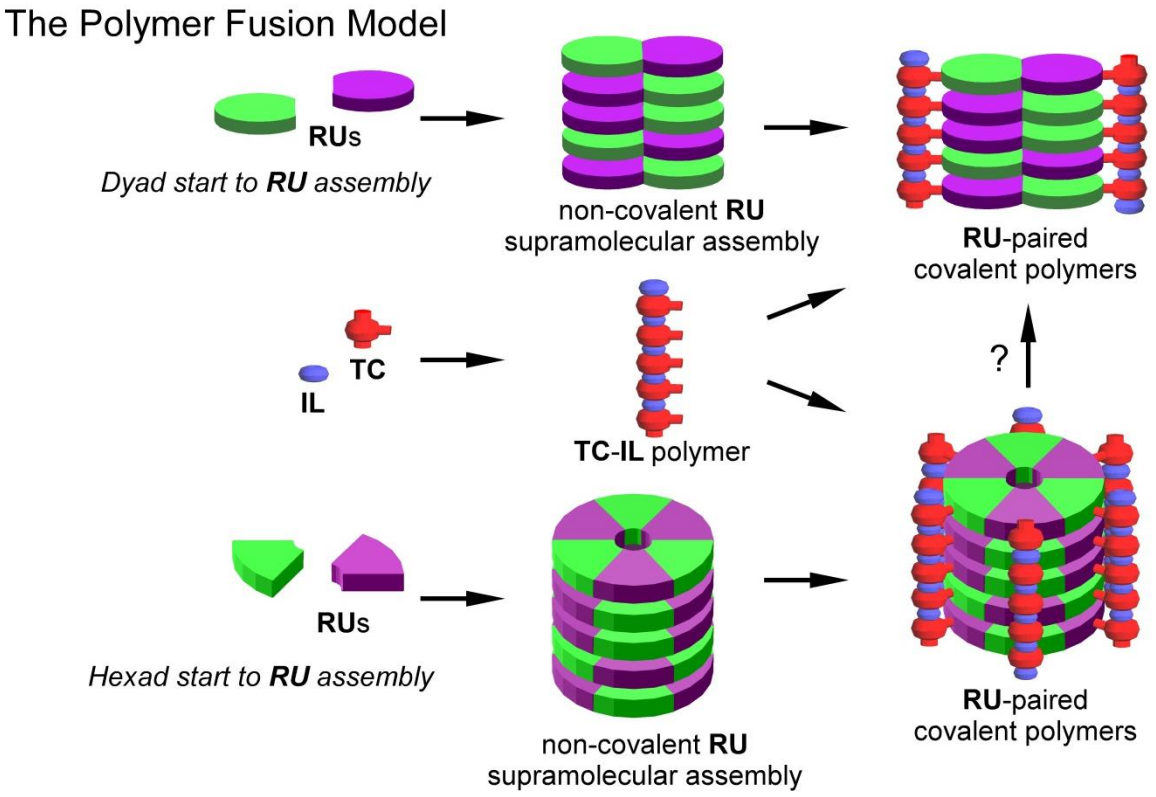


Figure 10.1 Schematic representation of the Polymer Fusion Model. Recognition units (**RU**s) form supramolecular assemblies that involve pairing, either as dyads or hexads, that are the same as those that will hold strands together in informational polymers. Trifunctional connector (TCs) and ionized linkers (ILs) form covalent polymers, among the many other polymers that exist in the prebiotic chemical inventory. The match in spacing of the functional groups of the TCs in the TC-IL polymers with the RUs in the supramolecular assemblies promotes the fusion of these polymers through the covalent linking of TCs and RUs.

10.2 Hypothetical Replication of Hexad-Based Proto-RNAs

The models and results discussed above provide a potential pathway for the *de novo* synthesis of a proto-RNA from self-assembling nucleobase analogs. In the chemical march towards life, abiotic replication of proto-RNA would, arguably, be the next great challenge that needed to be addressed. Hexad-mediated pairing between a preformed strand and complementary nucleotides is a feasible mechanism by which non-enzymatic template-directed polymerization could have transpired. That is, the polymerization of nucleosides organized within a hexad stack by one pre-formed strand of the stack would give rise to two identical strands and three strands with sequences complementary to the original template strand. Nucleobases with increased affinity for their complementary template are known to lead to higher yields and greater fidelity during nonenzymatic template directed RNA synthesis [7, 8]. It stands to reason that the hexad-based system which can assemble even as monomers and have more hydrogen bonding interactions, would be affected less by poor template/monomer affinity and mismatches. Thus, hexad-mediated template-directed synthesis may ultimately prove to be less prone to errors than reactions that rely solely on Watson-Crick base pairing between a template strand and monomers.

For replication to continue, the product strand(s) of a template-direct synthesis reaction must be separated in order for the original template and the new product strands to act as templates in subsequent reactions. Previous studies of duplex-mediated template-directed synthesis have demonstrated that the polymerization of free nucleosides (or the ligation of oligonucleotides) on a template strand can lead to a duplex of sufficient stability as to inhibit subsequent rounds of synthesis/replication, a quandary

often referred to in origins of life research as the strand inhibition problem [9]. Most models for circumventing the strand inhibition problem have suggested that intermittent high temperatures would have driven strand separation. In light of the greater affinity of TAP, BA and MA nucleosides for hexad stack formation, compared to Watson-Crick base pair formation by extant nucleobases, it might be expected that oligomers within a hexad assembly would be more challenging, or impossible, to separate in water. However as discussed in Chapter 3, we have demonstrated that the assembly of a TAP-based rosette system could be controlled isothermally with modest pH change around neutral pH [10]. This can be accomplished because the pK_a of TAP is 7, so when solution pH is below neutrality the base is predominantly positively charged and unable to enter the hydrophobic rosette assembly. In contrast the pK_a's of the natural system are around 4 (for C and A) and 9 (G and U) and would require significantly acidic or basic environments to enable pH-dependent disassembly, which would also greatly promote polymer degradation. Having pK_a's of pairing bases near neutrality enables slight, and less hostile, environmental pH variations to greatly influence monomer affinity. The ability to promote stand separation by changes in pH reduces the need to rely upon substantial increases in temperature for strand separation.

10.3 Towards Watson and Crick: Evolving from Pyrimidine/Triazine Proto-RNA

Siegel and Tor [11] have also proposed that RNA might have originated with pyrimidine-pyrimidine base pairs, and provided a conceptual path for the gradual evolution of putative original pyrimidine nucleobases to the extant nucleobases. In particular, these authors proposed that adenine could have evolved from the proto-nucleobase 2,4-diaminopyrimidine (**40**), which was connected to ribose through its

exocylic amine at the C4 position. This proposal is attractive, as such a pyrimidine nucleoside can, in principle, be converted to a purine nucleoside by the "closing" of a ring that contains the exocyclic amine and the C4, C5 carbons of the pyrimidine ring. Moreover, with ribose attached at the exocyclic amine of the pyrimidine or triazine, then the pairing of this nucleoside with another pyrimidine/triazine nucleoside (e.g., BA with ribose attached directly to the pyrimidine ring) would produce a base pair with a geometry that is extremely close to that of a Watson-Crick base pair. As discussed above, the exocyclic amines of TAP and MA readily react with ribose in water to form nucleosides that structurally mimic a diaminopyrimidine nucleoside, so the proposal of Siegel and Tor may ultimately be validated. Furthermore, consistent with the principle of continuity, the H-bonding face of BA is forward compatible with adenine, while TAP, possessing an additional H-bond donor, remains compatible with uracil recognition. In Figure 10.2 we present a hypothetical path in which, through discrete steps, a proto-RNA TAP•BA base pair evolves into the extant A•U base pair.

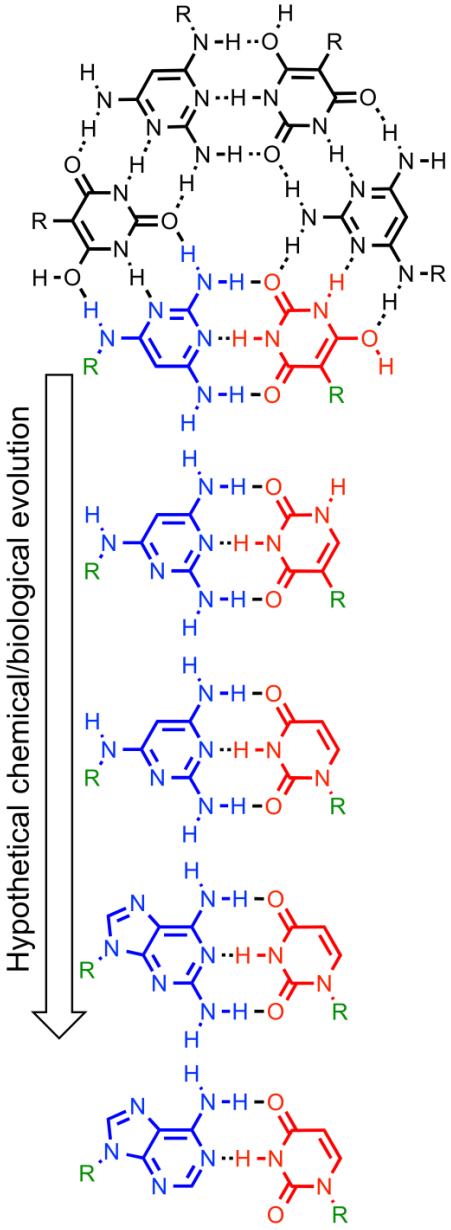


Figure 10.2 Hypothetical chemical and/or biological steps in the evolution of the proposed TAP•BA base pair of proto-RNA to an A•U base pair of extant RNA. Intermediate structures are not intended to suggest that evolutionary changes occurred in a specific order, or that there were not more or less steps over the course of evolution from proto-RNA to extant RNA. R indicates ribose or another trifunctional linker that may have predated ribose.

10.4 A Systematic Approach to the Origin of RNA

Based on the arguments and experimental results discussed throughout this thesis, the 91 heterocycles depicted in Figure 1.5 can be reduced to a considerably smaller number of candidate proto-nucleobases. As shown in Figure 10.3, removing the purines, pyrimidines, and triazines that would not be forward compatible with Watson-Crick base pairing (crossed out) and those that are known not to form nucleosides with sugars in model prebiotic reactions (crossed out with dashed lines) already reduces the number of candidate proto-RNA bases greatly. Emphasizing those bases that are known to form nucleosides with sugars in model prebiotic reactions (boxed in black), those that self-assemble in aqueous solution (in dashed boxes), and those that have been identified in meteorites or among the products of model prebiotic reactions (boxed in orange), clearly illustrate that TAP, BA, MA and Cy are currently our most attractive candidates for the original nucleobases of proto-RNA. It is certainly the case that more heterocycles will be demonstrated to fit our criteria of forming nucleosides in model prebiotic reactions, and more may also be demonstrated to self-assemble in aqueous solution. It is also likely that additional reasons will be identified to remove heterocycles from the list of possible candidates for the original bases of proto-RNA. For example, if molecular self-assembly is essential for candidacy, as we have discussed in Chapter 4, a number of heterocycles can be removed as they did not show the ability to be incorporated into a hexad assembly that, for most of them, would have maximized H-bonding. Going forward, the chemical landscape should be expanded beyond the purines, pyrimidines and s-triazines so as not to exclude the possibility that early informational polymers used other heterocycles. Alternatively, small aromatic molecules that promote base-paired assemblies of otherwise

non-pairing heterocycles and oligomers (see *Molecular Midwife Hypothesis*), should also be evaluated; several of the potential benefits for small molecule mediated pre-RNA assembly are discussed in Chapter 8.

In conclusion, defining a set of criteria to screen proto-nucleobase candidates, and then conducting experiments to test the ability for specific molecules to satisfy these criteria, provides a systematic way to potentially identify a set of molecules from which the original nucleic acids most likely emerged, as well as to identify informational systems potentially employed elsewhere in the universe. Towards this effort, several heterocycles have now been demonstrated to overcome multiple challenges that have long stood in the way of the spontaneous formation of RNA-like nucleotides and polymers. These same heterocycles have additionally provided insight into molecular self-assembly in water, including design principles to harness the hydrophobic effect and methods to parse the contribution of individual noncovalent interactions enabling the formation of large supramolecular structures. Future investigations will undoubtedly be needed to unravel additional properties of these and related self-assembling systems, which may be important for the development of new biomaterials, advances in synthetic biology and uncovering emergent properties during the transition from inanimate mater to life (e.g., mirror symmetry breaking via supramolecular assembly). These and other efforts will provide continued insight into the chemical etiology project originated by Albert Eschenmoser [12], and may eventually unveil a compelling chemical story for what was the structure of proto-RNA.

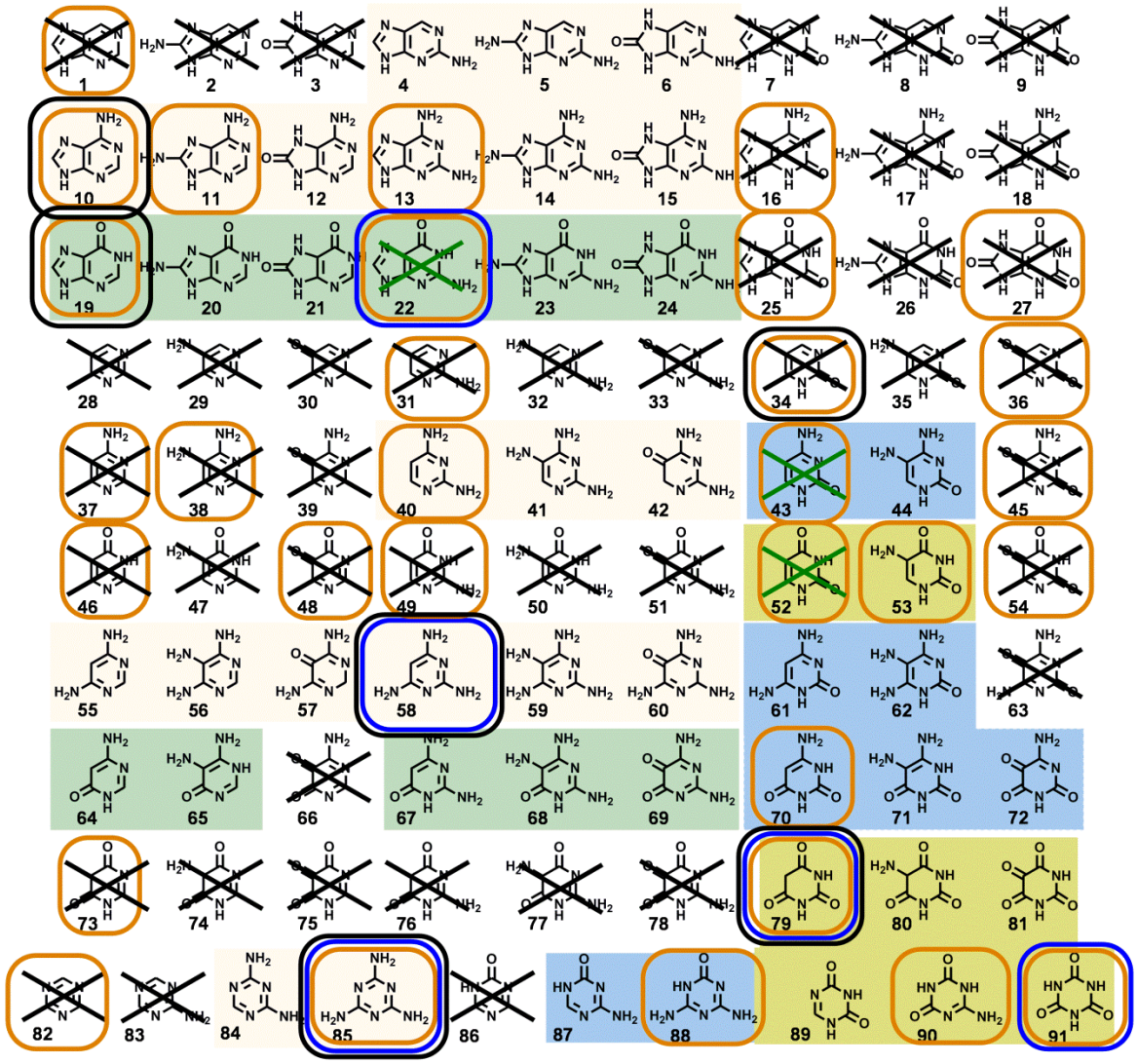


Figure 10.3 Evaluation of proto-nucleobase candidacy for the chemical space defined the purine, pyrimidine or triazine ring system and exocyclic groups containing O, N, or H (see Figure 1.5 for more details). Heterocycles that are not expected to form, as nucleosides, a Watson-Crick base pair with at least two H-bonds with an extant nucleobase (i.e., not forward compatible for information transfer) are crossed out. Heterocycles that are expected to be forward compatible are shown on colored backgrounds, yellow for pairing with adenine, cyan for pairing guanine, orange for pairing with uracil and green for pairing with cytosine. Heterocycles shown to form nucleosides with sugars are boxed in black [13-17], the three extant nucleobases that have not been demonstrated to form nucleosides with pre-formed sugars are crossed out with dashed lines. Heterocycles known to self-assemble as monomers in water are in blue boxes [18-20]. Heterocycles identified in meteorites [21, 22] or in model prebiotic reactions [23-30] are boxed in orange.

10.5 References

1. Albertazzi, L., et al., *Spatiotemporal control and superselectivity in supramolecular polymers using multivalency.* Proc Natl Acad Sci U S A, 2013. **110**: p. 12203-12208.

2. Safont-Sempere, M.M., G. Fernandez, and F. Wuerthner, *Self-Sorting Phenomena in Complex Supramolecular Systems.* Chem Rev, 2011. **111**: p. 5784-5814.

3. Sato, K., Y. Itoh, and T. Aida, *Homochiral supramolecular polymerization of bowl-shaped chiral macrocycles in solution.* Chem Sci, 2014. **5**: p. 136-140.

4. Hud, N.V., et al., *The origin of RNA and 'My Grandfather's Axe'.* Chem Biol, 2013. **20**: p. 466-474.

5. Lehn, J.-M., *Perspectives in Chemistry - Steps towards Complex Matter.* Angew Chem Int Ed, 2013. **52**: p. 2836-2850.

6. Li, J., P. Nowak, and S. Otto, *Dynamic Combinatorial Libraries: From Exploring Molecular Recognition to Systems Chemistry.* J Am Chem Soc, 2013. **135**: p. 9222-9239.

7. Zhang, S.L., et al., *Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer.* Proc Natl Acad Sci U S A, 2013. **110**: p. 17732-17737.

8. Hartel, C. and M.W. Gobel, *Substitution of adenine by purine-2,6-diamine improves the nonenzymatic oligomerization of ribonucleotides on templates containing thymidine.* Helvetica Chimica Acta, 2000. **83**: p. 2541-2549.

9. Fernando, C., G. Von Kiedrowski, and E. Szathmary, *A stochastic model of nonenzymatic nucleic acid replication: "Elongators" sequester replicators.* J Mol Evol, 2007. **64**: p. 572-585.

10. Cafferty, B.J., et al., *Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK(a) matching of biomimetic monomers.* Chem Sci, 2014. **5**: p. 4681-4686.

11. Siegel, J.S. and Y. Tor, *Genetic alphabetic order: what came before A?* Org Boimol Chem, 2005. **3**: p. 1591-1592.

12. Eschenmoser, A., *Chemical etiology of nucleic acid structure.* Science, 1999. **284**: p. 2118-2124.

13. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis. VI. synthesis of purine nucleosides.* J Mol Biol, 1972. **67**: p. 25-33.

14. Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis: VII. solid-state synthesis of purine nucleosides.* J Mol Evol, 1972. **1**: p. 249-257.

15. Bean, H.D., et al., *Formation of a β-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction.* J Am Chem Soc, 2007. **129**: p. 9556-9557

16. Chen, M.C., et al., *Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle.* J Am Chem Soc, 2014. **136**: p. 5640-5646.

17. Wulff, G. and G. Clarkson, *On the synthesis of C-glycosyl compounds containing double-bonds without the use of protecting groups.* Carbohydr Res, 1994. **257**: p. 81-95.

18. Davis, J.T., *G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry.* Angew Chem Int Ed Engl, 2004. **43**: p. 668-698.

19. Cafferty, B.J., et al., *Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues.* J Am Chem Soc, 2013. **135**: p. 2447-2450.

20. Bohanon, T.M., et al., *Barbituric-acid 2,4,6-triaminopyrimidine aggregates in water and their competitive interaction with a monolayer of barbituric-acid lipids at the gas-water interface.* Angew Chem Int Ed, 1995. **34**: p. 58-60.

21. Callahan, M.P., et al., *Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases.* Proc Natl Acad Sci U S A, 2011. **108**: p. 13995-13998.

22. Botta, O. and J.L. Bada, *Extraterrestrial organic compounds in meteorites.* Surv Geophys, 2002. **23**: p. 411-467.

23. Saladino, R., et al., *A possible prebiotic synthesis of purine, adenine, cytosine, and 4(3H)-pyrimidinone from formamide: Implications for the origin of life.* Bioorg Med Chem, 2001. **9**: p. 1249-1253.

24. Menor-Salvan, C., et al., *Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases.* Chem Eur J, 2009. **15**: p. 4411-4418.

25. Menor-Salvan, C. and M.R. Marin-Yaseli, *A new route for the prebiotic synthesis of nucleobases and hydantoins in water/ice solutions involving the photochemistry of acetylene.* Chem Eur J, 2013. **19**: p. 6488-6497.

26. Ferris, J.P. and W.J. Hagan, *HCN and chemical evolution: The possible role of cyano compounds in prebiotic synthesis.* Tetrahedron, 1984. **40**: p. 1093-1120.

27. Barks, H.L., et al., *Guanine, adenine, and hypoxanthine production in UV-irradiated formamide solutions: Relaxation of the requirements for prebiotic purine nucleobase formation.* ChemBioChem, 2010. **11**: p. 1240-1243.

28. Borquez, E., et al., *An investigation of prebiotic purine synthesis from the hydrolysis of HCN polymers.* Orig Life Evol Biosph, 2005. **35**: p. 79-90.

29. Cleaves, H.J., K.E. Nelson, and S.L. Miller, *The prebiotic synthesis of pyrimidines in frozen solution.* Naturwissenschaften, 2006. **93**: p. 228–231.

30. Nuevo, M., S.N. Milam, and S.A. Sandford, *Nucleobases and prebiotic molecules in organic residues produced from the ultraviolet photo-irradiation of pyrimidine in NH_3 and H_2O+NH_3 ices.* Astrobiology, 2012. **12**: p. 295-314.